

PREDIKSI HASIL PEMILU LEGISLATIF MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR BERBASIS *BACKWARD ELIMINATION*

Achmad Saiful Rizal¹, Moch. Lutfi²

^{1,2}Teknik Informatika, Universitas Yudharta Pasuruan
Jl. Yudharta No.7, Kembangkuning, Sengonagun, Purwosari, Pasuruan, Jawa Timur 67162, Indonesia

Email: cakaris99@gmail.com¹, moch.lutfi@yudharta.ac.id²

Received : Februari, 2020

Accepted : April, 2020

Published : April, 2020

Abstract

Elections in Indonesia from period to period have undergone some changes. Elections legislative candidates not determined voters, but instead became a political elite authority in accordance with the order of the list of legislative candidates and their number sequence. To perform a prediction one of them with data mining. Data mining can be applied in the political sphere for example to predict the results of the legislative election and others. K-nearest neighbor algorithm is one of the data mining algorithm that performs classification based on learning object against which are closest to the object. Election-related research has been done with the k-nearest neighbor algorithm, but accuracy is obtained that method is still too low, so it takes an additional algorithm to improve accuracy. In this study, the proposed method, namely the method of k-nearest neighbor method combined with backward elimination as a selection of features. The dataset that will be used in the study comes from the KPU Sidoarjo that has special attributes 1 and 13 regular attributes. From the results of the analysis and computation of some methods, it can be concluded that the method of k-nearest neighbor method combined with backward elimination produced some conclusions. First, of the 14 attributes in the dataset, retrieved 8 most influential attribute. Second, the best accuracy are of 96.03% when $k = 2$ and tested by 10 fold cross validation.

Keywords: Pemilu, K-Nearest Neighbor, Backward Elimination

Abstrak

Pemilu di Indonesia dari periode ke periode telah mengalami beberapa perubahan. Pemilihan calon legislatif tidak ditentukan pemilih, melainkan menjadi otoritas elit politik sesuai dengan urutan daftar calon legislatif dan urutan nomor mereka. Untuk melakukan prediksi salah satunya dengan data mining dapat diterapkan di bidang politik misalnya untuk memprediksi hasil pemilihan legislatif dan lainnya. Algoritma K-nearest neighbor adalah salah satu algoritma data mining yang melakukan klasifikasi berdasarkan pembelajaran objek terdekat. Penelitian terkait pemilu telah dilakukan dengan algoritma tetangga k-terdekat, tetapi akurasi yang diperoleh metode itu masih terlalu rendah, sehingga dibutuhkan algoritma tambahan untuk meningkatkan akurasi. Dalam penelitian ini, metode yang diusulkan, yaitu metode metode k-tetangga terdekat dikombinasikan dengan eliminasi mundur sebagai pilihan fitur. Dataset yang akan digunakan dalam penelitian ini berasal dari KPU Sidoarjo yang memiliki atribut khusus 1 dan 13 atribut reguler. Dari hasil analisis dan perhitungan beberapa metode, dapat disimpulkan bahwa metode metode k-tetangga terdekat dikombinasikan dengan backward elimination menghasilkan beberapa kesimpulan. Pertama, dari 14 atribut dalam dataset, diambil 8 atribut

yang paling berpengaruh. Kedua, akurasi terbaik adalah 96,03% ketika $k = 2$ dan diuji dengan validasi silang 10 kali lipat.

Kata kunci: Pemilu, *K-nearest neighbor*, *backward elimination*.

1. PENDAHULUAN

Pemilihan Umum di Indonesia dari periode ke periode telah mengalami beberapa perubahan. Sebagai contoh, sistem pemilu yang dilaksanakan pada Orde Baru, terkenal sebagai sistem pemilu yang tertutup. Keterpilihan calon legislatif bukan ditentukan pemilih, melainkan menjadi kewenangan elite politik sesuai dengan susunan daftar calon legislatif beserta nomor urut [1]. Banyak orang beranggapan bahwa dengan menjadi Dewan Perwakilan Rakyat orang dengan mudah mendapatkan uang banyak, sehingga banyak orang berlomba-lomba untuk mencalonkan diri. Karena jumlah kursi yang diperebutkan terbatas, maka banyak dari calon Dewan Perwakilan Rakyat yang tidak terpilih menjadi stress dikarenakan hartanya sudah habis untuk pencalonan dirinya. Sehingga dibutuhkan prediksi keterpilihan calon agar dapat mengukur dan memantapkan calon tersebut.

Untuk melakukan sebuah prediksi salah satu satunya yaitu dengan data mining. Data mining dapat diaplikasikan di bidang politik misalnya untuk memprediksi hasil pemilu legislatif dan lain-lain. Terdapat beberapa metode dalam memprediksi hasil pemilu legislatif misalnya C4.5 [2], *Naïve Bayes* [3], *Neural Network* [4], *K-Nearest Neighbor* [5] dan lain-lain. Menurut Moradian dan Barani [6] algoritma *K-Nearest Neighbor* adalah salah satu algoritma klasifikasi yang digunakan sebagian besar dalam aplikasi yang berbeda. Salah satu masalah dari algoritma *K-Nearest Neighbor* adalah semua atribut dalam menghitung jarak antara *record* baru dan *record* yang tersedia dalam dataset training. Hal ini menyebabkan proses klasifikasi yang tidak baik dan menurunkan akurasi algoritma klasifikasi. Dalam pembahasan ini, terbukti bahwa akurasi

yang didapatkan ketika prediksi hasil pemilu legislatif menggunakan algoritma *K-Nearest Neighbor* menghasilkan akurasi rendah seperti hasil penelitian yang telah dilakukan oleh Muhammad Badrul [5] yang menghasilkan akurasi sebesar 83.35%. Berbeda dengan penelitian yang dilakukan oleh Evicienna dan Hilda Amalia [2] menggunakan algoritma C4.5 yang menghasilkan akurasi sebesar 97.84%, Diana Tri Wahyudi, T. Sutojo, dan Ardytha Luthfiarta [3] menggunakan algoritma *Naïve Bayes* dan algoritma genetika yang menghasilkan akurasi sebesar 97.84%, dan Muhammad Badrul [4] menggunakan algoritma *Neural Network* yang menghasilkan akurasi sebesar 98.50%.

Backward Elimination adalah metode seleksi fitur yang terbukti dapat meningkatkan akurasi pada algoritma klasifikasi. Seperti pada penelitian yang dilakukan oleh Laily Hermawati [7], bahwa metode *Backward Elimination* dapat meningkatkan akurasi metode *Naïve Bayes* dengan dataset penyakit kanker payudara, semula menghasilkan akurasi sebesar 96.14% menjadi 97.00%, serta penelitian yang dilakukan oleh Laily Hermawati dan Sucianna Ghadati Rabiha [8], bahwa metode *Backward Elimination* dapat meningkatkan akurasi metode *K-Nearest Neighbor* dengan dataset penyakit jantung, semula menghasilkan akurasi sebesar 88.62% menjadi 89.55%. Pada penelitian ini, peneliti akan menggunakan metode *k-Nearest Neighbor* dengan *backward elimination* sebagai seleksi fitur untuk memprediksi hasil pemilu legislatif Kabupaten Sidoarjo pada tahun 2014.

Penelitian tentang prediksi hasil pemilu sudah banyak dilakukan. Antara lain penelitian yang dilakukan oleh Evicienna dan Hilda Amalia [2] melakukan

penerapan algoritma C4.5 pada data Pemilu legislatif DKI Jakarta tahun 2009 dengan 11 atribut menghasilkan akurasi sebesar 97.84%. Penelitian yang sama dilakukan oleh Diana Tri Wahyuni, T. Sutojo, dan Ardytha Luthfiarta [3] melakukan penerapan algoritma *Naïve Bayes* dengan algoritma genetika sebagai fitur seleksi pada data Pemilu legislatif DKI Jakarta tahun 2009 menghasilkan akurasi sebesar 97.84%. Penelitian yang dilakukan oleh Mohammad Badrul [4] melakukan penerapan algoritma *Neural Network* pada data Pemilu legislatif DKI Jakarta tahun 2009 menghasilkan akurasi sebesar 98.10% dan nilai AUC sebesar 0.987. Setelah dilakukan perubahan pada beberapa parameter yang meliputi *training cycle*, *learning rate*, *momentum*, *neuron size* dan *hidden layer* diperoleh nilai akurasi baru yaitu 98,50 % dan nilai AUC sebesar 0.982.

Kemudian penelitian yang dilakukan oleh Alfin Duhawan Bagja, Gunawan Abdullah dan Faiza Renaldi [9] melakukan penerapan algoritma *K-Nearest Neighbor* pada data Pemilu legislatif daerah pemilihan Jawa Barat dengan menentukan daerah pemilihan yang berpotensi berhasil bagi bakal calon legislatif berdasarkan 14 atribut yang menghasilkan akurasi sebesar 85.62%. Terakhir, penelitian yang dilakukan oleh Mohammad Badrul [5] melakukan penerapan algoritma *K-Nearest Neighbor* pada data Pemilu legislatif DKI Jakarta tahun 2009 menghasilkan akurasi sebesar 81.35%. Dalam penelitian ini ada 6 atribut yang membedakan dengan dataset Pemilu legislatif DKI Jakarta pada tahun 2009, antara lain: pendidikan terakhir, suara sah partai dan caleg per-dapil, total suara sah seluruh partai dan caleg per-dapil, alokasi kursi per-dapil, total suara sah partai dan caleg, dan peringkat suara sah caleg per-dapil.

2. LANDASAN TEORI

Data mining merupakan ilmu yang mempelajari cara untuk mengekstrak dan mengeksplorasi suatu pengetahuan atau

menemukan suatu pola dari sebuah data dengan menggunakan beberapa metode yang sesuai. Data mining merupakan salah satu cara untuk menemukan informasi yang ada dalam sebuah database dan merupakan bagian dari proses *Knowledge Discovery in Database* (KDD). Rangkaian proses atau tahap data mining dapat dibagi menjadi beberapa bagian [10], yaitu:

1. Pembersihan data, yaitu untuk membuang data yang tidak konsisten dan noise.
2. Integrasi data, yaitu penggabungan data dari beberapa sumber.
3. Transformasi data, yaitu data diubah dan diolah menjadi bentuk yang bisa atau yang sesuai untuk di-mining atau proses pengolahan data berikutnya.
4. Penerapan metode, yaitu melakukan pengolahan data dengan metode yang dipilih.
5. Evaluasi hasil, yaitu untuk mengetahui kinerja dari metode yang digunakan.
6. Presentasi pengetahuan yaitu membuat pelaporan eksperimen secara keseluruhan.

Konversi data merupakan proses awal dalam pemrosesan dataset. Proses ini diperlukan agar dataset bisa diolah sesuai dengan kebutuhan. Konversi data nominal menjadi data numerik dilakukan dengan menetapkan ranking *value* yang akan di konvert. Proses penetapan ranking data nominal tidak sama dengan data numerik. Jika pada data numerik langsung dengan melihat *value* yang ada, maka pada data nominal harus melalui proses. Salah satunya dengan cara menghitung frekwensi data nominal [11] yang akan diubah menjadi numerik.

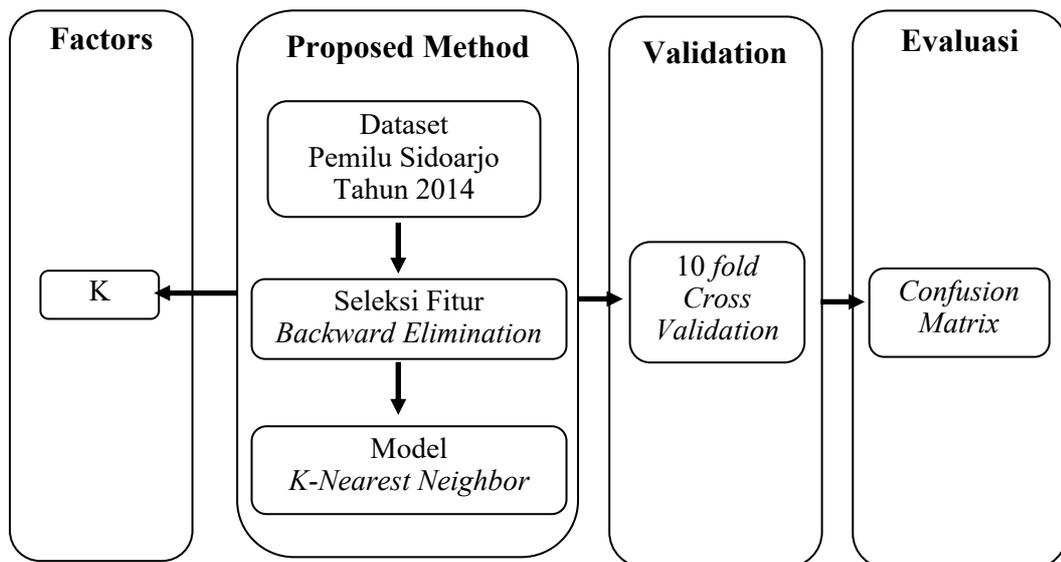
Seleksi fitur bertujuan untuk membuat model dataset agar didapatkan model terbaik bagi pengolahan data selanjutnya. Sedangkan *backward elimination* [12] merupakan metode untuk menghapus atribut yang tidak relevan. Prinsip *backward elimination* adalah memasukan semua

variabel bebas yang ada, kemudian mengeliminasi variabel satu persatu.

Algoritma *K-Nearest Neighbor* [13] merupakan metode klasifikasi objek berdasarkan dataset training yang jaraknya berdekatan dengan objek itu. Dataset training diklasifikasikan ke dalam ruang yang mempunyai dimensi yang banyak, yangmana setiap dimensi mewakili ciri khas dari data setiap dimensi. Sebuah titik di dalam ruang ditandai dengan kelas n jika kelas n adalah yang paling banyak ditemui pada sejumlah k tetangga terdekat dari titik itu. Persamaan jarak antar tetangga dihitung dengan jarak Euclidian:

$$d(x-y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} = \sqrt{\sum_{j=1}^n (x_j - y_j)^2} \quad (1)$$

2.1. Kerangka Penelitian



Gambar 2.1 Kerangka Pemikiran

Pada kerangka diatas menggambarkan alur penyelesaian masalah pada dataset pemilu legislatif sidoarjo, yang mana data di proses dengan seleksi fitur dengan menggunakan metode backward elimination selanjutnya data di proses dengan algoritma k-nearest neighbor dengan menentukan parameter K pada algoritma tersebut.

Keterangan:

d = jarak data uji ke dataset training

x_j = data uji ke- j dimana $j=1,2,\dots,n$

y_j = dataset training ke- j , dimana $j = 1,2,\dots,n$.

Uji t-statistik [14] atau t-hitung merupakan pengujian untuk mengetahui apakah masing- masing koefisien regresi signifikan atau tidak terhadap variabel dependen. Dengan menganggap variabel independen lainnya konstan. Nilai t-hitung diperoleh dengan rumus:

$$t\text{-hitung} = (b_i - b) / S_e(b_i) \quad (2)$$

Dimana;

b_i = koefisien variabel ke- i

b = nilai hipotesis nol

$S_e(b_i)$ = simpangan baku dari variabel independen ke- i

3. METODE PENELITIAN

3.1. Pengumpulan Data

Data dalam penelitian ini adalah data pemilu legislatif yang diperoleh dari database KPU Kabupaten Sidoarjo pada tahun 2014. Dataset ini terdiri dari 528 data dengan 14 atribut.

Tabel 3.1 dataset pemilu

Nama Partai	Nomer Urut Caleg	Nama Caleg	Jk	Pendidikan Terakhir	Daerah Pemilihan	Suara Sah Caleg	Suara Sah Partai	Suara Sah Partai dan Caleg per-Dapil	Total Suara Sah Seluruh Partai dan Caleg per-Dapil	Alokasi Kursi per-Dapil	Total Suara Sah Partai dan Caleg	Jumlah Perolehan Kursi per-Dapil	Peringkat Suara Sah Caleg per-Dapil	Terpilih atau Tidak
Partai Nasdem	1	Drs. AKHMAD NADJIB MARTAK	L	S1	1	1218	1387	6588	209160	11	38036	0	2	Tidak
Partai Nasdem	2	H. MOH. MARSAM IBAD, S.Ag.	L	S1	1	418	1387	6588	209160	11	38036	0	4	Tidak
Partai Nasdem	3	MASRIFATUS SHOLIKHAH, S.T.	P	S1	1	230	1387	6588	209160	11	38036	0	7	Tidak
...
Partai Nasdem	4	SAIFUL BACHRI	L	SLTA	1	1640	1387	6588	209160	11	38036	0	1	Tidak

3.2. Konversi Data

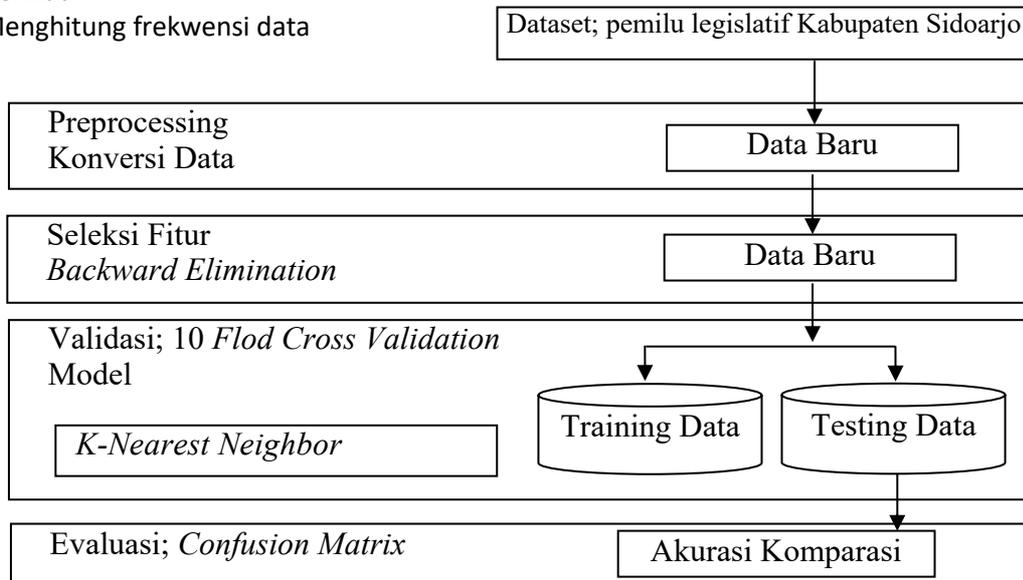
Dataset yang digunakan dalam penelitian ini memiliki atribut campuran, yaitu polinomial dan numerik. Maka dari itu data perlu di terjemahkan ke dalam format numerik agar memudahkan dalam pengolahan data secara keseluruhan. Proses konversi dilakukan dengan cara sebagai berikut:

1. Menghitung frekwensi data

2. Menetapkan ranking dari tiap data berdasarkan nilai frekwensi

3.3. Metode yang diusulkan

Metode yang diusulkan yaitu metode dengan penerapan algoritma *K-Nearest Neighbor* berbasis *backward elimination* sebagai seleksi fitur.



Gambar 3.1. Metode yang diusulkan

Dataset pemilu legislatif merupakan dataset private yang di gunakan dalam penelitian

ini, sebelum data di learning dengan metode yang diusulkan data terlebih dahulu

di preprocessing data atau di konversi dan di seleksi fitur dengan metode backward yang mana tahap preprocessing ini digunakan untuk memastikan data benar-benar sudah bisa di learning dengan algoritma k-nearest neighbor untuk menentukan performance yang terbaik dari metode yang diusulkan.

Seleksi Fitur dengan *Backward Elimination*

Seleksi fitur ini bertujuan untuk mencari model terbaik dari dataset. Langkah yang bisa dilakukan yaitu dengan melakukan regresi terhadap tiap atribut di dalam dataset. Adapun tahapan dalam metode *backward elimination* adalah sebagai berikut:

1. Membuat model pengolahan data dengan memasukkan semua atribut.
2. Kemudian membuat perbandingan pengolahan data dengan mengeluarkan salah satu atribut ke-n.
3. Kemudian diulangi hingga semua atribut n+1.
4. Kemudian semua hasil perhitungan dibandingkan hingga didapatkan hasil atribut manakah yang relevan dan manakah yang irrelevant.

Klasifikasi dengan Metode *K-Nearest Neighbor*

Setelah proses *backward elimination* selesai, maka langkah berikutnya adalah mengklasifikasikan dataset dengan metode *K-Nearest Neighbor*. Adapun tahapannya sebagai berikut:

1. Menentukan regular atribut sebagai label
2. Menentukan K terdekat
3. Menghitung jarak terdekat

Validasi dengan *Cross Validation*

Merupakan sebuah metode validasi yang menggunakan semua data yang ada menjadi *training set* dan *test set*. Dalam metode ini, dibentuk k *subset* dari dataset yang ada. Contohnya, *10-fold cross validation* yang berarti bahwa 9 *subset* menjadi *training set* dan 1 *subset* sebagai *testing set*, dengan 10 kali pengulangan pengukuran [15]. Adapun tahapan *cross*

validation pada penelitian ini sebagai berikut:

1. Memecah 528 dataset ke dalam 10 bagian, yaitu 9 bagian dataset training sebanyak 90% atau 475 dan 1 bagian dataset testing sebanyak 10% atau 53.
2. Menghitung jarak. Kasus awal dilakukan pada data ke-1 sebagai data training dan data ke-476 sebagai data testing. Dan kemudian menghitung kedekatan jarak antar nilai atribut dari data training dan testing.
3. Dan seterusnya hingga data training ke-9.

Evaluasi dengan *Confusion Matrix*

Merupakan tabel klasifikasi yang mengandung informasi hasil perhitungan sistem secara keseluruhan[16]. Pengukuran data dievaluasi melalui akurasi, presisi dan *recall*. Hasil pengukuran direpresentasikan ke dalam sebuah tabel klasifikasi untuk memudahkan pembacaan. Tahapan evaluasi dengan *Confusion matrix* adalah:

1. Menentukan Nilai *True Positive* (TP) dan nilai *False Negative* (FN).
2. Menentukan persentase *class recall* dari masing-masing kolom *true* Terpilih dan Tidak Terpilih
3. Menentukan persentase *class precision* dari masing-masing baris prediksi Terpilih dan Tidak Terpilih
4. Menentukan persentase *accuracy* secara keseluruhan

Uji beda statistik dengan *t-test*

Membandingkan hasil dari setiap metode apakah memiliki perbedaan yang signifikan atau tidak dengan tahapan sebagai berikut:

1. Menghitung kuadrat selisih
2. Menentukan standar deviasi
3. Menentukan t hitung
4. Menentukan t tabel
5. Membandingkan t hitung dan t tabel
6. Menarik kesimpulan

4. HASIL PENELITIAN DAN PEMBAHASAN

4.1. Pengumpulan Data

Data dalam penelitian ini diperoleh dari KPU Kabupaten Sidoarjo tahun 2014 yang terdiri

dari 528 data dengan 14 atribut. Berikut keterangan atribut pada dataset.

1. Nomor Urut Partai sebagai *regular attribute*.
2. Nomor Urut Caleg sebagai *regular attribute*.
3. Jenis Kelamin sebagai *regular attribute*.
4. Pendidikan Terakhir sebagai *regular attribute*.
5. Daerah Pemilihan sebagai *regular attribute*.
6. Suara Sah Caleg sebagai *regular attribute*.
7. Suara Sah Partai sebagai *regular attribute*.
8. Suara Sah Partai dan Caleg per-Dapil sebagai *regular attribute*.
9. Total Suara Sah Seluruh Partai dan Caleg per-Dapil sebagai *regular attribute*.
10. Alokasi Kursi per-Dapil sebagai *regular attribute*.
11. Total Suara Sah Partai dan Caleg sebagai *regular attribute*.
12. Jumlah Perolehan Kursi per-Dapil sebagai *regular attribute*.
13. Peringkat Suara Sah Caleg per-Dapil sebagai *regular attribute*.
14. Terpilih atau Tidak sebagai *special attribute* atau sebagai *label*

4.2. Konversi Data

Tahapan konversi adalah dengan metode menghitung nilai frekwensi pada setiap *value*. Kemudian menentukan nilai ranking berdasarkan nilai frekwensi. Adapun hasil dari proses konversi sebagaimana berikut:

1. Nomor Urut Partai; 1 (1), 2 (2), 3 (3), 4 (4), 5 (5), 6 (6), 10 (7), 8 (8), 7 (9), 15 (10), 14 (11), 9 (12)
2. Nomor Urut Caleg; 1 (1), 2 (2), 3 (3), 4 (4), 5 (5), 6 (6), 7 (7), 8 (8), 9 (9), 10 (10), 11 (11)
3. Jenis Kelamin; L (1) dan P (2)
4. Pendidikan Terakhir; Sarjana (1) dan SLTA (2)
5. Daerah Pemilihan; 1 (1), 5 (2), 6 (3), 4 (4), 2 (5), 3 (6)
6. Suara Sah Caleg; *numeric*
7. Suara Sah Partai; *numeric*
8. Suara Sah Partai dan Caleg per-Dapil; *numeric*
9. Total Suara Sah Seluruh Partai dan Caleg per-Dapil; *numeric*
10. Alokasi Kursi per-Dapil; *numeric*
11. Total Suara Sah Partai dan Caleg; *numeric*
12. Jumlah Perolehan Kursi per-Dapil; *numeric*
13. Peringkat Suara Sah Caleg per-Dapil; *numeric*
14. Terpilih atau Tidak; *binominal* (Terpilih dan Tidak) sebagai label

4.3. Seleksi Fitur dengan **Backward Elimination**

Seleksi fitur dengan *backward elimination* bertujuan untuk mengeliminasi atribut yang tidak berpengaruh signifikan pada dataset. Adapun langkah-langkah yang dilakukan melakukan perhitungan *backward elimination* adalah:

1. Membuat model pengolahan data dengan memasukkan semua atribut. Terlebih dahulu adalah menentukan regresi atribut X dan Y

Tabel 4.1 Menentukan X dan Y

Nomor Urut Partai	Nomor Urut Caleg	Jk	Pendidikan Terakhir	Daerah Pemilihan	Suara Sah Caleg	Suara Sah Partai	Suara Sah Partai dan Caleg per-Dapil
X1	X2	X3	X4	X5	X6	X7	X8

Tabel lanjutan

Total Suara Sah Seluruh Partai	Alokasi Kursi per-Dapil	Total Suara Sah Partai dan Caleg	Jumlah Perolehan Kursi per-Dapil	Peringkat Suara Sah Caleg	Terpilih atau Tidak
--------------------------------	-------------------------	----------------------------------	----------------------------------	---------------------------	---------------------

dan Caleg per-Dapil				per-Dapil	
X9	X10	X11	X12	X13	Y

Tabel di atas merupakan penentuan masing-masing atribut untuk keperluan memasukkan variabel ke dalam persamaan regresi sebagai berikut:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8 + b_9X_9 + b_{10}X_{10} + b_{11}X_{11} + b_{12}X_{12} + b_{13}X_{13} \quad (3)$$

2. Kemudian membuat perbandingan pengolahan data dengan mengeluarkan salah satu atribut ke-n.

Setelah melakukan perhitungan regresi, atribut yang mempunyai nilai *P-value* terbesar dari nilai *P-value Intercept*, maka akan dieliminasi karena dianggap atribut yang tidak berpengaruh atau tidak signifikan. Kemudian langkah selanjutnya yaitu mengulangi langkah ke-1 dan ke-2

dengan meregresi seluruh atribut sampai ditemukan nilai *P-value* pada setiap atribut tidak lebih besar dari nilai *P-value Intercept*, maka proses perhitungan akan selesai dan dapat diperoleh atribut yang terbaik.

4.4. Pembobotan pada Atribut

Pembobotan pada atribut ini dimaksudkan untuk mengetahui atribut mana yang mempunyai pengaruh terbesar hingga terkecil pada prediksi keterpilihan calon legislatif. Adapun tahapan-tahapannya sebagaimana berikut:

1. Menentukan ranking tiap atribut
Penentuan ranking ini bersifat subyektif [17] yang dipengaruhi oleh persepsi pengambil keputusan. Dalam penelitian ini penentuan ranking berdasarkan besar nilai *P-value*.

Tabel 4.2 Rangkaing atribut

No	Atribut	P-value	Ranking
1	Nomor Urut Partai	0.82843	9
2	Nomor Urut Caleg	0.87378	10
3	Jenis Kelamin	0.79290	8
4	Pendidikan Terakhir	0.41700	5
5	Daerah Pemilihan	0.98974	13
6	Suara Sah Caleg	0.00000	1
7	Suara Sah Partai	0.09746	4
8	Suara Sah Partai dan Caleg per-Dapil	0.00003	3
9	Total Suara Sah Seluruh Partai dan Caleg per-Dapil	0.90464	12
10	Alokasi Kursi per-Dapil	0.70992	6
11	Total Suara Sah Partai dan Caleg	0.78943	7
12	Jumlah Perolehan Kursi per-Dapil	0.00001	2
13	Peringkat Suara Sah Caleg per-Dapil	0.87956	11

Dari tabel di atas bisa ditentukan bahwa dari semua atribut dataset yang memiliki nilai *P-value* terkecil adalah suara sah caleg, sehingga variabel suara sah caleg memiliki ranking pertama.

2. Menentukan bobot tiap atribut

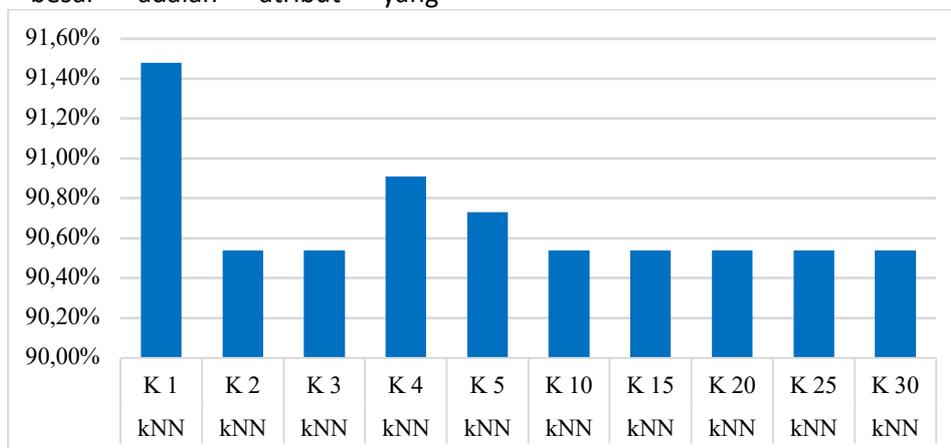
$$w_j = \frac{(n-r_j+1)}{\sum(n-r_p+1)} \quad (4)$$

Dari perhitungan dengan rumus di atas, maka diperoleh ranking dan bobot tiap atribut dari dataset yang digunakan.

Tabel 4.3 Bobot atribut

No	Atribut	Ranking	Bobot
1	Suara Sah Caleg	1	0.143
2	Jumlah Perolehan Kursi per-Dapil	2	0.132
3	Suara Sah Partai dan Caleg per-Dapil	3	0.121
4	Suara Sah Partai	4	0.111
5	Pendidikan Terakhir	5	0.099
6	Alokasi Kursi per-Dapil	6	0.088
7	Total Suara Sah Partai dan Caleg	7	0.077
8	Jenis Kelamin	8	0.066
9	Nomor Urut Partai	9	0.055
10	Nomor Urut Caleg	10	0.044
11	Peringkat Suara Sah Caleg per-Dapil	11	0.033
12	Total Suara Sah Seluruh Partai dan Caleg per-Dapil	12	0.022
13	Daerah Pemilihan	13	0.011

Tabel di atas, atribut dengan bobot yang paling besar adalah atribut yang



Gambar 4.1 Grafik hasil perhitungan metode kNN dengan 10 fold cross validation

mempunyai pengaruh signifikan pada proses perhitungan komputasi dengan menggunakan algoritma *K-Nearest Neighbor*.

4.5. Klasifikasi menggunakan Metode *K-Nearest Neighbor*

Adapun tahapan dalam metode *K-Nearest Neighbor* yaitu:

1. Menentukan regular atribut sebagai label
2. Menentukan K terdekat
Pada contoh ini, peneliti bereksperimen dengan menentukan K=1 s.d. 30
3. Menghitung jarak terdekat dengan menggunakan rumus *euclidean distance*
Pada eksperimen pertama, peneliti menggunakan 10 *fold cross validation*. Kemudian hasil perhitungan tersebut ditunjukkan pada tabel dan grafik berikut.

Tabel 4.4 Hasil perhitungan metode kNN dengan 10 fold cross validation

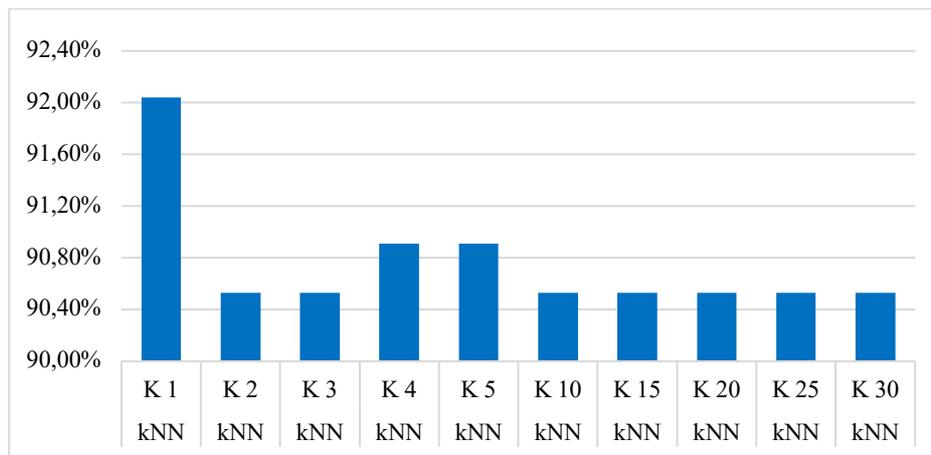
K	Akurasi
1	91.48% +/-2.56%
2	90.54% +/-1.42%
3	90.54% +/-1.42%
4	90.91% +/-1.09%
5	90.73% +/-1.28%
10	90.54% +/-1.15%
15	90.54% +/-1.15%
20	90.54% +/-1.15%
25	90.54% +/-1.15%
30	90.54% +/-1.15%

Pada eksperimen kedua, peneliti menggunakan 5 *fold cross validation*. Kemudian hasil perhitungan tersebut ditunjukkan pada tabel dan grafik berikut.

Tabel 4.5 Hasil perhitungan metode kNN dengan 5 *fold cross validation*

K	Akurasi
1	92.04% +/-0.47%
2	90.53% +/-0.63%

3	90.53% +/-0.63%
4	90.91% +/-0.46%
5	90.91% +/-0.46%
10	90.53% +/-0.57%
15	90.53% +/-0.57%
20	90.53% +/-0.57%
25	90.53% +/-0.57%
30	90.53% +/-0.57%



Gambar 4.2 Grafik hasil perhitungan metode kNN dengan 5 *fold cross validation*

Dari hasil kedua eksperimen tersebut dapat disimpulkan bahwa, akurasi terbaik metode *K-Nearest Neighbor* yaitu sebesar 92.04%. Dari hasil tersebut, Peneliti merasa masih perlu untuk meningkatkan akurasi tersebut. Oleh karena itu dibutuhkan metode tambahan untuk meningkatkan akurasi tersebut.

Untuk meningkatkan hasil akurasi metode *K-Nearest Neighbor*, peneliti bereksperimen dengan menggabungkan antara metode *K-Nearest Neighbor* dengan metode *backward elimination* sebagai fitur seleksi. Pada eksperimen pertama, peneliti menggunakan 10 *fold cross validation*. Kemudian hasil perhitungan tersebut ditunjukkan pada tabel dan grafik berikut.

4.6. Klasifikasi Metode *K-Nearest Neighbor* dengan *Backward Elimination*

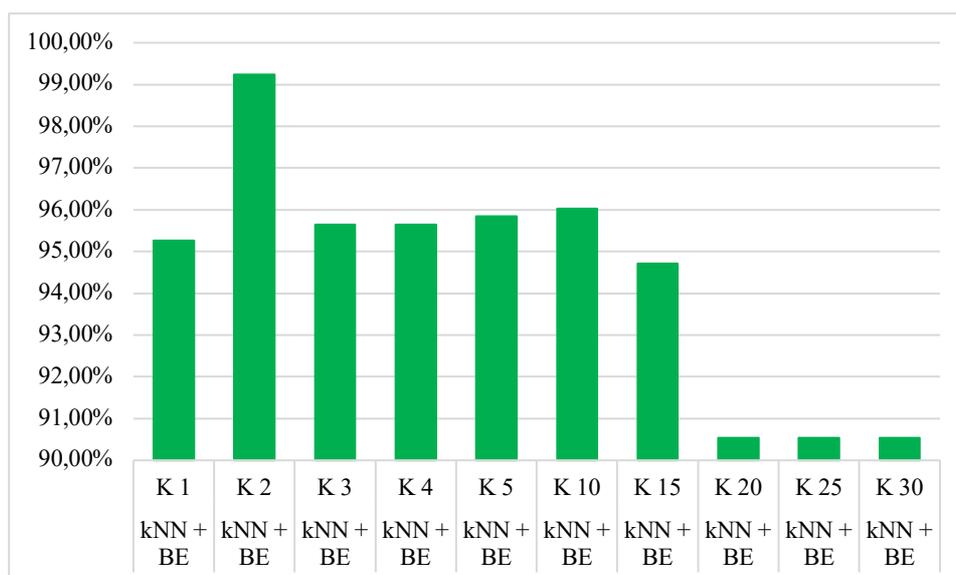
Tabel 4.6 Hasil penggabungan perhitungan metode kNN dengan metode *backward elimination* menggunakan 10 *fold cross validation*

K	Atribut yang terpilih	Akurasi
1	Nomer Urut Partai, Daerah pemilihan, Suara Sah Caleg, Alokasi Kursi per-Dapil, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	95.27% +/-2.11%
2	Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	99.24% +/-1.73%
3	Nomer Urut Partai, Nomer Urut Caleg, Pendidikan Terakhir, Daerah pemilihan, Suara Sah Caleg, Alokasi Kursi per-Dapil, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	95.65% +/-1.20%

4	Nomer Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah pemilihan, Suara Sah Caleg, Alokasi Kursi per-Dapil, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	95.65% +/-2.07%
5	Nomer Urut Partai, Nomer Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah pemilihan, Suara Sah Caleg, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	95.84% +/-1.84%
10	Nomer Urut Partai, Nomer Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah pemilihan, Suara Sah Caleg, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	96.03% +/-1.56%
15	Nomer Urut Partai, Nomer Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah pemilihan, Suara Sah Caleg, Suara Sah Partai, Alokasi Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	94.71% +/-3.02%
20	Nomor Urut Partai, Nomor Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah Pemilihan, Suara Sah Caleg, Suara Sah Partai, Suara Sah Partai dan Caleg per-Dapil, Total Suara Sah Seluruh Partai dan Caleg per-Dapil, Alokasi Kursi per-Dapil, Total Suara Sah Partai dan Caleg, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	90.54% +/-1.15%

Tabel lanjutan

K	Atribut yang terpilih	Akurasi
25	Nomor Urut Partai, Nomor Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah Pemilihan, Suara Sah Caleg, Suara Sah Partai, Suara Sah Partai dan Caleg per-Dapil, Total Suara Sah Seluruh Partai dan Caleg per-Dapil, Alokasi Kursi per-Dapil, Total Suara Sah Partai dan Caleg, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	90.54% +/-1.15%
30	Nomor Urut Partai, Nomor Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah Pemilihan, Suara Sah Caleg, Suara Sah Partai, Suara Sah Partai dan Caleg per-Dapil, Total Suara Sah Seluruh Partai dan Caleg per-Dapil, Alokasi Kursi per-Dapil, Total Suara Sah Partai dan Caleg, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	90.54% +/-1.15%



Gambar 4.3 Grafik hasil penggabungan perhitungan metode kNN dengan metode *backward elimination* menggunakan 10 *fold cross validation*

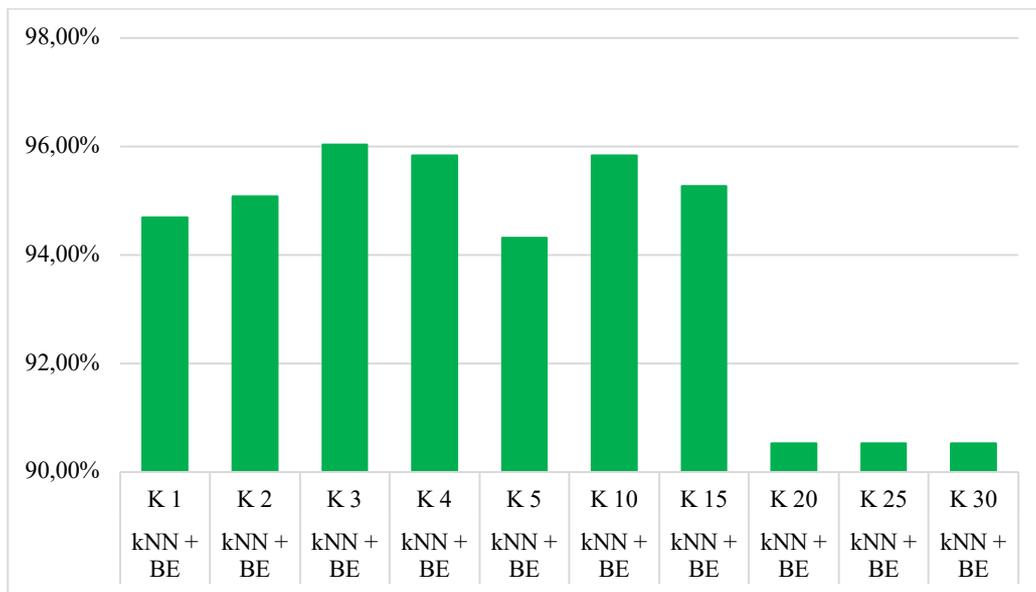
Pada eksperimen pertama, peneliti menggunakan *5 fold cross validation*. Kemudian hasil perhitungan tersebut ditunjukkan pada tabel dan grafik berikut.

Tabel 4.7 Hasil penggabungan perhitungan metode kNN dengan metode *backward elimination* menggunakan *5 fold cross validation*

K	Atribut yang terpilih	Akurasi
1	Nomor Urut Partai, Nomor Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah Pemilihan, Suara Sah Caleg, Suara Sah Partai, Suara Sah Partai dan Caleg per-Dapil, Alokasi Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	94.69% +/-1.56%
2	Nomor Urut Partai, Suara Sah Caleg, Alokasi Kursi per-Dapil, Total Suara Sah Partai dan Caleg, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	95.08% +/-0.69%
3	Nomor Urut Partai, Nomor Urut Caleg, Suara Sah Caleg, Alokasi Kursi per-Dapil, Total Suara Sah Partai dan Caleg, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	96.03% +/-1.83%

Tabel lanjutan

K	Atribut yang terpilih	Akurasi
4	Nomor Urut Partai, Nomor Urut Caleg, Pendidikan Terakhir, Daerah Pemilihan, Suara Sah Caleg, Alokasi Kursi per-Dapil, Jumlah Perolehan Kursi per-Dapil	95.84% +/-0.95%
5	Nomor Urut Partai, Nomor Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah Pemilihan, Suara Sah Caleg, Suara Sah Partai, Alokasi Kursi per-Dapil, Total Suara Sah Partai dan Caleg, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	94.32% +/-1.19%
10	Nomor Urut Partai, Jenis Kelamin, Pendidikan Terakhir, Daerah Pemilihan, Suara Sah Caleg, Alokasi Kursi per-Dapil, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	95.83% +/-1.41%
15	Nomor Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah Pemilihan, Suara Sah Caleg, Suara Sah Partai, Jumlah Perolehan Kursi per-Dapil	95.27% +/-1.98%
20	Nomor Urut Partai, Nomor Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah Pemilihan, Suara Sah Caleg, Suara Sah Partai, Suara Sah Partai dan Caleg per-Dapil, Total Suara Sah Seluruh Partai dan Caleg per-Dapil, Alokasi Kursi per-Dapil, Total Suara Sah Partai dan Caleg, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	90.53% +/-0.57%
25	Nomor Urut Partai, Nomor Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah Pemilihan, Suara Sah Caleg, Suara Sah Partai, Suara Sah Partai dan Caleg per-Dapil, Total Suara Sah Seluruh Partai dan Caleg per-Dapil, Alokasi Kursi per-Dapil, Total Suara Sah Partai dan Caleg, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	90.53% +/-0.57%
30	Nomor Urut Partai, Nomor Urut Caleg, Jenis Kelamin, Pendidikan Terakhir, Daerah Pemilihan, Suara Sah Caleg, Suara Sah Partai, Suara Sah Partai dan Caleg per-Dapil, Total Suara Sah Seluruh Partai dan Caleg per-Dapil, Alokasi Kursi per-Dapil, Total Suara Sah Partai dan Caleg, Jumlah Perolehan Kursi per-Dapil, Peringkat Suara Sah Caleg per-Dapil	90.53% +/-0.57%



Gambar 4.4 Grafik hasil penggabungan perhitungan metode kNN dengan metode *backward elimination* menggunakan *5 fold cross validation*

Dari hasil kedua eksperimen tersebut, akurasi terbaik metode *K-Nearest Neighbor* yang digabung metode *Backward Elimination* yaitu sebesar 99.24%. Ini menunjukkan bahwa, metode *Backward Elimination* dapat meningkatkan performa metode *K-Nearest Neighbor*.

4.7. Evaluasi Hasil Penelitian dengan *Confusion matrix*

Tabel 4.8 *Confusion matrix backward elimination* dan kNN

	<i>True Tidak</i> (0)	<i>True Terpilih</i> (1)	<i>Class Precision</i>
<i>Pred. Tidak</i> (0)	475	18	96.35%
<i>Pred. Terpilih</i> (1)	3	32	91.43%
<i>Class Recall</i>	99.37%	64.00%	

Dari tabel hasil di atas, hasil akurasi perhitungan metode *backward elimination* dan *K-Nearest Neighbor* dengan $K=10$ dan *10 fold cross validation* adalah sebagai berikut:

- Nilai *True Positive* untuk Tidak Terpilih adalah 475 dan Terpilih adalah 32. Sedangkan nilai *False Negative* untuk Tidak Terpilih adalah 3 dan Terpilih adalah 18.
- Class recall* merupakan kolom yang berisi besar nilai klasifikasi yang tepat. Misalnya *class recall* dari

Pada penelitian ini evaluasi yang digunakan untuk mengetahui tingkat akurasi perhitungan dengan metode *backward elimination* dan *K-Nearest Neighbor* dengan $K=10$ dan menggunakan *10 fold cross validation* yaitu menggunakan *confusion matrix*.

Tidak Terpilih yang tepat pengklasifikasiannya adalah:

$$recall = \frac{TP}{TP + FN} = \frac{475}{475 + 3} = \frac{475}{478} = 0.9937 = 99.37\%$$

- Class precision* merupakan baris yang berisi besar nilai klasifikasi yang tepat. Misalnya *class precision* dari Tidak Terpilih yang tepat pengklasifikasiannya adalah:

$$precision = \frac{TP}{TP + FN} = \frac{475}{475 + 18} = \frac{475}{493} = 0.9635$$

$$= 96.35\%$$

d. *Accuracy* merupakan persentase hasil klasifikasi yang benar.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + FN} =$$

$$\frac{475 + 32}{478 + 3 + 18 + 32} = \frac{507}{528} = 0.9603$$

$$= 96.03\%$$

4.8. Uji Beda (*t*-test)

Secara keseluruhan, hasil percobaan metode *K-Nearest Neighbor* dan penambahan metode *backward elimination* didapatkan persentase akurasi sebagaimana berikut.

Tabel 4.9 Hasil akurasi

Metode	K=1	K=2	K=3	K=4	K=5	K=10	K=15	K=20	K=25	K=30
kNN	91.48	90.54	90.54	90.91	90.73	90.54	90.54	90.54	90.54	90.54
kNN + BE	92.27	99.24	95.65	95.65	95.84	96.03	94.71	90.54	90.54	90.54

Tabel di atas merupakan nilai akurasi dari dua metode yaitu *K-Nearest Neighbor* dan *K-Nearest Neighbor* yang dikombinasikan *backward elimination*. Nilai akurasi dari dua metode tersebut digunakan sebagai acuan untuk menentukan kuadrat selisih deviasi.

Tabel 4.10 Deviasi (selisih)

K	kNN (x1)	kNN + BE (x2)	D = x1-x2	D ²
1	91.48	92.27	-0.79	0.6241
2	90.54	99.24	-8.70	75.6900
3	90.54	95.65	-5.11	26.1121
4	90.91	95.65	-4.74	22.4676
5	90.73	95.84	-5.11	26.1121
10	90.54	96.03	-5.49	30.1401
15	90.54	94.71	-4.17	17.3889
20	90.54	90.54	0.00	0.0000
25	90.54	90.54	0.00	0.0000
30	90.54	90.54	0.00	0.0000
	Jumlah		-34.11	198.5349

Kemudian menentukan standar deviasi dengan cara sebagai berikut:

$$s = \sqrt{\frac{1}{n-1} \left\{ \sum D^2 - \frac{(\sum D)^2}{n} \right\}}$$

$$= \sqrt{\frac{1}{10-1} \left\{ 198.5349 - \frac{(-34.11)^2}{10} \right\}} =$$

$$\sqrt{\frac{1}{9} \left\{ 198.5349 - \frac{1163.4921}{10} \right\}}$$

$$= \sqrt{0.1111 \{198.5349 - 116.3492\}} =$$

$$\sqrt{0.1111 \{82.1857\}}$$

$$= \sqrt{9.1317} = 3.0219$$

Kemudian menghitung *t* hitung dengan cara sebagai berikut:

$$t \text{ hitung} = \frac{\frac{-34.11}{10}}{\frac{3.0219}{\sqrt{10}}} = \frac{-3.411}{0.956}$$

$$= -3.569$$

Kemudian menghitung *t* tabel dengan cara sebagai berikut:

Dk pembilang = variabel bebas = x = 2
 Dk penyebut = jumlah data – variabel bebas - 1 = n-k-1 = 10-2-1= 7

Dk total = Dk pembilang + Dk penyebut = 2 + 7 = 9

t tabel = 1-TDIST(*t* hitung, Dk pembilang, 1) = 0.971405

t hitung < *t* tabel, sehingga H₀ ditolak dan H_a diterima yang artinya penambahan *backward elimination* pada metode *K-Nearest Neighbor* memiliki perbedaan yang signifikan.

4.9. Membandingkan Akurasi Hasil Pengujian

Setelah dilakukan eksperimen, maka tahap terakhir yaitu membandingkan akurasi antara kedua metode.

Tabel 4.11 Perbandingan akurasi antara metode kNN dengan metode kNN yang digabung *backward elimination*

Metode	K	Akurasi
kNN	1	91.48% +/-2.56%
kNN	2	90.54% +/-1.42%
kNN	3	90.54% +/-1.42%
kNN	4	90.91% +/-1.09%
kNN	5	90.73% +/-1.28%
kNN	10	90.54% +/-1.15%
kNN	15	90.54% +/-1.15%
kNN	20	90.54% +/-1.15%
kNN	25	90.54% +/-1.15%
kNN	30	90.54% +/-1.15%
kNN + BE	1	95.27% +/-2.11%
kNN + BE	2	99.24% +/-1.73%
kNN + BE	3	95.65% +/-1.20%
kNN + BE	4	95.65% +/-2.07%
kNN + BE	5	95.84% +/-1.84%
kNN + BE	10	96.03% +/-1.56%
kNN + BE	15	94.71% +/-3.02%
kNN + BE	20	90.54% +/-1.15%
kNN + BE	25	90.54% +/-1.15%
kNN + BE	30	90.54% +/-1.15%

Dari perbandingan tabel di atas, metode *backward elimination* dapat meningkatkan akurasi dengan sangat baik. Sebagai contoh, ketika metode *K-Nearest Neighbor* dengan $K=1$ menghasilkan akurasi sebesar 91.48% meningkat menjadi 95.27% ketika digabung dengan metode *backward elimination*. Sedangkan akurasi terbaik pada perbandingan antara kedua metode tersebut sebesar 99.24% yaitu ketika metode *K-Nearest Neighbor* dengan *backward elimination* menggunakan $K=2$ dan menggunakan 10 *fold cross validation*.

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Dari hasil analisa diatas dapat disimpulkan bahwa metode *K-Nearest Neighbor* berbasis *backward elimination* untuk prediksi hasil pemilu legislatif menghasilkan:

1. Dari 14 atribut pada dataset pemilu legislatif, maka diperoleh 8 atribut yang paling berpengaruh terhadap prediksi hasil pemilu legislatif, yaitu atribut nomer urut partai, nomer urut caleg, jenis kelamin, pendidikan terakhir, daerah pemilihan, suara sah caleg,

jumlah perolehan kursi per-dapil, dan peringkat suara sah caleg per-dapil.

2. Penggunaan metode *K-Nearest Neighbor* yang digabungkan dengan metode *backward elimination* pada dataset pemilu legislatif dengan 14 atribut dan 528 record data, terbukti dapat meningkatkan akurasi. Sebagai contoh ketika $K=10$ dan diuji dengan 10 *fold cross validation* perhitungan metode *K-Nearest Neighbor* menghasilkan akurasi sebesar 90.54%, sedangkan metode *K-Nearest Neighbor* yang digabungkan dengan metode *backward elimination* menghasilkan akurasi sebesar 96.03%.

DAFTAR PUSTAKA

- [1] DPR RI, "Undang-Undang Republik Indonesia Nomor 10," 2008.
- [2] Eviciana and H. Amalia, "Algoritma C4.5 Untuk Prediksi Hasil Pemilihan Legislatif DPRD DKI Jakarta," *Thecno Nusa Mandiri*, vol. IX, no. 1, pp. 48–56, 2013.
- [3] D. T. Wahyuni, T. Sutojo, and A. Luthfiarta, "Prediksi Hasil Pemilu Legislatif DKI Jakarta Menggunakan Naïve Bayes Dengan Algoritma Genetika Sebagai Fitur Seleksi," *UDINUS*, 2004.
- [4] M. Badrul, "Penerapan Metode Neural Network Untuk Memprediksi Hasil Pemilu Legislatif," *Techno Nusa Mandiri*, vol. X, no. 2, pp. 21–32, 2013.
- [5] M. Badrul, "Prediksi Hasil Pemilu Legislatif Dengan Menggunakan Algoritma k-Nearest Neighbor," *Pilar Nusa Mandiri*, vol. XI, no. 2, pp. 152–160, 2015.
- [6] M. Moradian and A. Baraani, "Knnba : K-Nearest-Neighbor-Based-Association Algorithm," *Theor. Appl. Inf. Technol.*, 2009.
- [7] L. Hermawati, "Penggabungan Algoritma Backward Elimination dan Naive Bayes Untuk Mendiagnosis Penyakit Kanker Payudara," *Momentum*, vol. 11, no. 1, pp. 42–45, 2015.

- [8] L. Hermawati and S. G. Rabiha, "Penggabungan Algoritma Backward Elimination dan k-Nearest Neighbor Untuk Mendiagnosis Penyakit Jantung," *Pros. SNST*, pp. 184–189, 2014.
- [9] A. D. Bagja, G. Abdillah, and F. Renaldi, "Penerapan Data Mining Dalam Menentukan Potensi Keberhasilan Bakal Calon Legislatif di Daerah Pemilihan Jawa Barat Menggunakan Algoritma k-Nearest Neighbors," *Pros. SNST*, pp. 184–189, 2016.
- [10] J. I. Luke and Suharjito, "Data Mining Tweet Promosi Produk dan Jasa Secara Otomatis dengan Menggunakan Algoritma Naive Bayes untuk Meningkatkan Engagement Followers Twitter," *Binus Univ.*, 2015.
- [11] Z. Gniazdowski and M. Grabowski, "Numerical Coding of Nominal Data," *Zesz. Nauk. WWSI*, vol. 9, no. 12, pp. 53–61, 2015.
- [12] R. S. Wardani and Purwanto, "Model Pengambilan Keputusan Dalam Prediksi Kasus Tuberkulosis Menggunakan Regresi Logistik Berbasis Backward Elimination," *Univ. Muhammadiyah Semarang*, 2014.
- [13] D. Aryanie and Y. Heryadi, "American Sign Language-Based Finger-Spelling Recognition Using k-Nearest Neighbors Classifier," *Int. Conf. Inf. Commun. Technol.*, pp. 533–536, 2015, doi: 10.1109/ICoICT.2015.7231481.
- [14] Sudjana, *Statika Untuk Ekonomi dan Niaga*, vol. 2. Tarsito, Bandung, 1993.
- [15] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *Int. Jt. Conf. Artif. Intell.*, vol. 14, no. 12, pp. 1137–1143, 1995, doi: 10.1067/mod.2000.109031.
- [16] Moch. Lutfi and Mochamad Hasyim, "Penanganan Data Missing Value Pada Kualitas Produksi Jagung Dengan Menggunakan Metode K-Nn Imputation Pada Algoritma C4.5," *J. Resist. (Rekayasa Sist. Komputer)*, vol. 2, no. 2, pp. 89–104, 2019, doi: 10.31598/jurnalresistor.v2i2.427.
- [17] M. B. Selamat, "Pembobotan Parameter dan Penentuan Keputusan," *Jur. Ilmu Kelaut. FIKP UH*, vol. IV, pp. 38–49, 2002.