

JPTK



Jurnal Pendidikan Teknologi dan Kejuruan



Volume 15, Nomor 1, Januari 2018

Jurnal Pendidikan Teknologi dan Kejuruan
Fakultas Teknik dan Kejuruan
Universitas Pendidikan Ganesha

Hal. 1 – 156

**ISSN (Print)
0216-3241**

**ISSN (Online)
2541-0652**

<http://ejournal.undiksha.ac.id/index.php/JPTK>

OPTIMASI ALGORITMA GENETIKA PADA K-NN UNTUK MEMPREDIKSI KECENDERONGAN “BLOG POSTING”

Hani Harafani

Teknik Informatika, STMIK Nusa Mandiri Jakarta
haniharafani@gmail.com

Abstrak

Menentukan faktor utama pola kecenderungan postingan blog merupakan hal yang penting. Dataset BLOGGER diambil dari situs UCI Machine Learning Repositories yang mana merupakan dataset dari Iran tepatnya provinsi Kohkiluyeh dan Boyer Ahmad. Dalam penelitian ini peneliti memilih metode k-NN karena k-NN dapat mengatasi masalah klasifikasi dengan baik, namun dibalik keunggulan k-NN pengaturan nilai k yang sangat sensitif kerap menjadi kelemahan dari k-NN. Algoritma Genetika pada penelitian ini telah berhasil memperbaiki akurasi prediksi k-NN dengan memberikan nilai k yang optimal bagi k-NN dengan hasil akurasi sebesar 88% dibandingkan dengan akurasi k-NN saja sebesar 82% dan metode yang digunakan pada penelitian sebelumnya sebesar 82%. Bahkan k-NN+ GA terbukti lebih unggul dari pada metode prediksi lainnya. Tujuan dari penelitian ini adalah menerapkan Algoritma Genetika untuk mendapatkan nilai k yang optimal sehingga dapat memberikan prediksi yang lebih akurat.

Kata kunci: Blog Posting, k-NN+GA, Algoritma Genetika

Abstract

Determine the main factor of bloggers tendency is the important thing. BLOGGER dataset is taken from UCI Machine Learning Repositories site which taken from Kohkiluyeh and Boyer Ahmad Province in Iran. In this research, researcher choose the k-NN method because it solve the problem of classification well, but the sensitivity of tuning k value often becomes a drawback. Genetic Algorithm in this research has been successful in improving the accuracy of k-NN prediction by providing the optimum k value for k-NN method with 88% accuracy compared with just k-NN in 82% accuracy and the previous method in 82% accuracy too. Even, k-NN + GA proved to be superior to other predictive method. The goal of this research is to apply Genetic Algorithm to get the optimal k value so as to give more accurate prediction.

Keywords : Blog Posting, k-NN+GA, Genetic Algorithm

PENDAHULUAN

Situs jejaring sosial telah menjadi bagian integral dari kehidupan kita sehari-hari (Ullah et al., 2017), dengan sebagian besar populasi dunia berinteraksi menggunakan berbagai situs jejaring sosial seperti Facebook, LinkedIn, dan Twitter. Blog adalah salah satu alat efektif (Gharehchopogh, 2012) yang dianggap sebagai salah satu modul utama dan kemampuan sosial dan interaktif dalam mewujudkan dunia teknologi informasi yang indah untuk kehidupan maya dan dunia virtual. Anggota blog yang bisa membuat postingan blog dikenal sebagai blogger, atau penulis blog (Gharehchopogh, Khaze, & Maleki, 2015). Secara umum, blog dikelola oleh satu orang tapi bisa dikelola oleh sekelompok

penulis. Kolam semua web log dari web sosial disebut "blogosphere". Blogging merupakan cara yang populer bagi orang untuk mengekspresikan pandangan mereka tentang berbagai masalah sosial (Ullah et al., 2017). Pandangan yang diungkapkan bisa dalam bentuk teks atau multimedia, dan posting ini sudah diketahui sebagai "posting blog". Karena kemampuan blog yang luas, mengibaratkan blog sebagai buku harian online yang mudah diakses dan tidak dipungut biaya (Gharehchopogh et al., 2015). Bukan hanya sebagai diary saja, namun Pengguna bisa mendapatkan umpan balik yang tepat dan gagasan terkait tentang artikel yang mereka kirim atau hadir dalam diskusi sosial dan politik dengan yang lain melalui blog mereka. Pembicaraan dan komunikasi virtual antar

pengguna melalui blog menunjukkan virtual interaksi dan kerjasama yang berakibat semakin populernya blog di seluruh dunia. Sehingga saat ini, kita bukan hanya konsumen informasi saja pada saat kita menggunakan web untuk mendapatkan informasi (Ullah et al., 2017), tapi juga produsen informasi karena kita dapat berbagi informasi menggunakan saluran web sosial.

Ada proporsi tertentu antara fitur blog dan kecenderungan postingan blogger diantaranya dengan pola sosial, politik dan budaya dari berbagai negara dan negara. Di bidang politik, biasanya para Blogger bekerja dalam kolaborasi erat. Mereka menciptakan kelompok kepentingan khusus mereka sendiri di mana mereka memulai diskusi sosial atau politik, menyebarkan kesadaran dan mengumpulkan dukungan (Ullah et al., 2017). Misalnya, mereka dapat mempengaruhi opini publik dengan membantu mengorganisir dan memobilisasi kampanye sosial dan politik.

Menyadari penyebab kecenderungan postingan para blogger dan parameter utama pendekatan mereka adalah di antara isu utama dimana perencanaan makro negara ditentukan berdasarkan teknologi modern dan penggunaannya dan memberikan hal yang penting (Gharehchopogh, 2012). Jadi, penting untuk memberikan solusi tepat untuk menentukan faktor utama pola kecenderungan postingan blog.

Beberapa metode machine learning banyak digunakan oleh peneliti dunia untuk mengklasifikasi pola kecenderungan postingan blog, diantaranya: Metode c4.5 (Gharehchopogh, 2012) dengan accuracy 82%, Random Tree, RandomForest (Asim, Shahid, Malik, & Raza, 2017), dan masih banyak lagi. Namun pada penelitian ini penulis tertarik untuk menggunakan k-NN sebagai metode untuk memprediksi kecenderungan dalam membuat postingan blog. Namun, metode k-NN sering sensitif terhadap pemilihan nilai k yang sesuai (Zhang, Cheng, Deng, Zong, & Deng, 2017).

Algoritma metaheuristik banyak diaplikasikan untuk mencari nilai yang optimal seperti Algoritma Genetika

(Harafani & Wahono, 2015), *Particle Swarm Optimization* (Harish, Mandal, Rao, & Patil, 2015). Dalam penelitian ini peneliti ingin mencoba peruntungan pada Algoritma Genetika sebagai algoritma yang populer dalam menemukan nilai yang optimal.

K- NEAREST NEIGHBOUR (K-NN)

k-NN (*k- Nearest Neighbour*) dikenal juga sebagai "*Lazy Learner*" (Setiawan, 2017). k-NN merupakan perpanjangan dari metode pengklasifikasi *Nearest Neighbour* (*NN classifier*) (Souza, Rittner, & Lotufo, 2014). Kerangka k-NN pertama kali diperkenalkan oleh Fix and Hodges (1951, 1952). Meskipun k-NN secara formal diajukan lebih dari 60 tahun yang lalu, namun k-NN masih tetap populer hingga saat ini, karena banyak sekali penelitian yang menggunakan metode ini, sebagai contohnya adalah mendeteksi kanker paru-paru (Bhuvaneswari & Therese, 2015), Klasifikasi karakter melalui tulisan tangan (Wakahara & Yamashita, 2014), Klasifikasi Data Besar (Zhang, Deng, Cheng, & Zhu, 2015), dan masih banyak lagi. Metode ini membandingkan kesamaan catatan pelatihan yang paling dekat dengannya. Nilai k mewakili tetangga yang dibandingkan.

```

start
  select a positive integer k along with new samples
  select the k entries in the database is closed to the sample
  find the most common classification of these entries
  this is the classification that will be given to the new sample
end
  
```

Gambar 1. Pseudocode k-NN Classifier

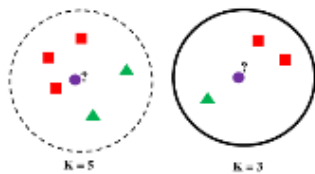
Kesamaan dihitung berdasarkan jarak, seperti jarak euclidean. Jarak antara dua nilai, misalnya X_1 dan X_2 , dimana $X_1 = (x_{11}, x_{12}, x_{13}, \dots, x_{1n})$ dan $X_2 = (x_{21}, x_{22}, x_{23}, \dots, x_{2n})$ didefinisikan oleh:

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^n (X_{1i} - X_{2i})^2} \quad (1)$$

Normalisasi min-max adalah yang diterapkan, untuk mengubah nilai v dari atribut numerik A ke v 'di kisaran 0 Ini didefinisikan oleh:

$$v' = \frac{v - \text{Min}_A}{\text{Max}_A - \text{Min}_A} \quad (2)$$

Dalam pengenalan pola, algoritma K-Nearest Neighbor (K-NN) adalah metode non-parametrik (Bhuvaneswari & Therese, 2015) yang digunakan untuk klasifikasi dan regresi. Dalam K-NN Classification, output adalah keanggotaan kelas. Klasifikasi dilakukan dengan suara terbanyak dari tetangga. Jika $K = 1$, maka kelasnya adalah satu tetangga terdekat seperti pada Gambar 2.



Gambar 2. Contoh Klasifikasi dengan k-NN

ALGORITMA GENETIKA

Algoritma genetika merupakan algoritma evolusioner yang paling populer (Yang, 2014) yang mana algoritma ini menggunakan prinsip dasar dari seleksi alam yang diperkenalkan oleh Charles Darwin. Algoritma genetika diterapkan sebagai pendekatan untuk mengidentifikasi pencarian nilai dan solusi bagi berbagai permasalahan optimasi (Gorunescu, 2011). Algoritma genetika memiliki tiga operator genetika utama yaitu *crossover* (proses penukaran kromosom), mutasi (proses penggantian salah satu solusi untuk meningkatkan keragaman populasi), seleksi (penggunaan solusi dengan nilai fitness yang tinggi untuk lulus ke generasi berikutnya). Terdapat langkah-langkah yang sering dilakukan untuk menyelesaikan permasalahan-permasalahan dalam optimasi:

1. Inisialisasi populasi
2. Evaluasi populasi
3. Seleksi populasi
4. Proses penyilangan kromosom (*crossover*)
5. Evaluasi populasi baru
6. Selama syarat belum terpenuhi

ulangi dari langkah 3.

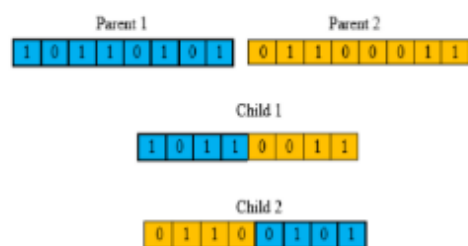
Terdapat beberapa kelebihan dari algoritma genetika dibandingkan algoritma optimasi tradisional lainnya, dua diantaranya yaitu kemampuan untuk menangani permasalahan kompleks dan parallel. Algoritma genetika dapat menangani berbagai macam optimasi tergantung pada fungsi objektifnya (*fitness*) apakah seimbang atau tidak seimbang, linier atau tidak linier, berkesinambungan atau tak berkesinambungan, atau dengan *random noise*. Fungsi Fitness (Kuncheva, 1997) ditunjukkan pada Persamaan(3).

$$J(S) = \sum_{j=1}^n hs(Jz) \quad (3)$$

dimana $hs(Z)$ menyatakan jumlah yang Z_j berkontribusi terhadap penilaian keseluruhan kriteria, diberikan S , dan menggunakan aturan klasifikasi k-NN.

Intisari dari algoritma genetika mencakup pengkodean/penyandian fungsi optimasi sebagai *array* berisi bit-bit atau karakter berupa sting untuk menggambarkan kromosom, operasi manipulasi string dengan operator genetika, dan seleksi sesuai dengan *fitness*, dengan tujuan untuk menemukan solusi yang baik dan optimal terhadap masalah yang sedang dihadapi. Algoritma genetika memiliki tiga operator genetika utama yaitu:

1. *Crossover* yaitu proses menukar bagian dari solusi (kromosom) dengan bagian "*parent*" lain untuk menghasilkan jenis kromosom yang berbeda yang mungkin menjadi solusi baru untuk menyelesaikan permasalahan. Peran utamanya adalah untuk memberikan pencampuran solusi dan konvergensi dalam sub ruang (menghasilkan solusi yang baru).



Gambar 3. Proses Cross Over pada Algoritma Genetika

2. Mutasi yaitu pergantian salah satu bagian solusi yang dipilih secara acak, yang meningkatkan keragaman dari populasi dan menghasilkan mekanisme untuk menghindari minimum lokal.



Gambar 4. Proses Mutasi pada Algoritma Genetika

3. Seleksi *fittest* atau elitisme yaitu penggunaan solusi dengan nilai *fitness* yang tinggi untuk lulus ke generasi berikutnya, yang sering dilakukan dalam hal beberapa bentuk seleksi dari solusi terbaik.

Menurut Haupt dalam (Zukhri,2014) terdapat langkah-langkah yang sering dilakukan untuk menyelesaikan permasalahan-permasalahan dalam optimasi:

1. Inisialisasi populasi.
2. Evaluasi populasi.
3. Seleksi populasi.
4. Proses penyilangan pasangan kromosom tertentu (*crossover*).
5. Evaluasi populasi baru.

Selama syarat belum terpenuhi ulangi dari langkah 3.

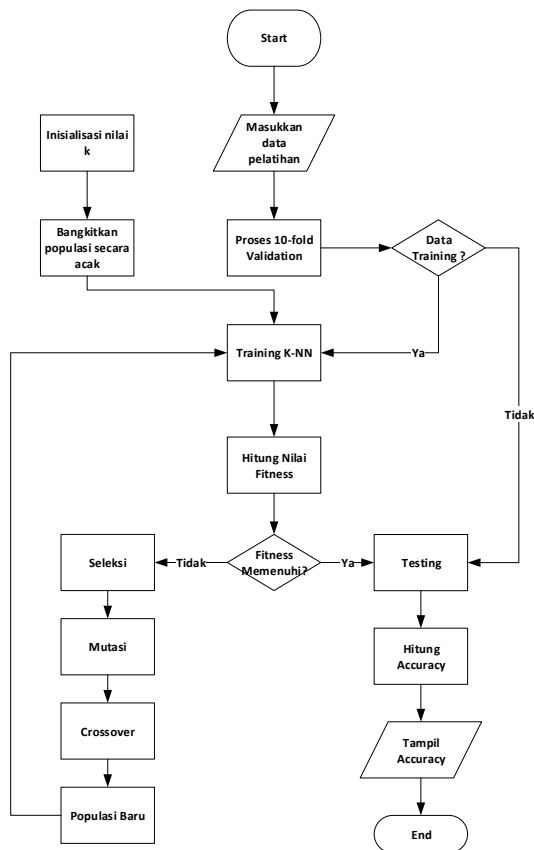
```

Fungsi Objektif  $f(x), x = (x_1, \dots, x_d)^T$ 
Menyandikan solusi menjadi kromosom (string)
Definisikan fitness  $F(eg, F \propto f(x)$  untuk maksimasi)
Menghasilkan populasi awal
Inisialisasi kemungkinan crossover ( $p_c$ ) dan mutasi ( $p_m$ )
While ( $t < \text{Jumlah maksimal generasi}$ )
    Hasilkan solusi baru dengan crossover dan mutasi
    Crossover dengan kemungkinan crossover ( $p_c$ )
    Mutasi dengan kemungkinan mutasi ( $p_m$ )
    Menerima semua solusi jika fitness meningkat
    Pilih hasil yang terbaik untuk generasi selanjutnya (elitisme)
End while
Mengembalikan hasil dan visualisasi ke bentuk awal (decode)
    
```

Gambar 5. Pseudocode Penyelesaian Masalah pada GA

METODE

Tahapan penelitian dimulai dari pengumpulan dataset. Dataset BLOGGER peneliti dapat dari <https://archive.ics.uci.edu/ml/datasets/BLOGGER> yang mana merupakan dataset dari Iran tepatnya provinsi Kohgiluyeh dan Boyer Ahmad. Awalnya (Gharehchopogh, 2012) mengumpulkan data dengan menyebar questioner lewat website yang kemudian diisi oleh para blogger di Iran. Hasilnya beberapa parameter yang dianggap sebagai bidang yang utama didapatkan. Diantaranya: *education, political caprice, topics, Local Media Turnover* (LMT), *local, political and social space* (LPSS).



Gambar 6. Metode Penelitian yang diusulkan

Dari seratus *record*, untuk mendapatkan jawaban yang benar, (Gharehchopogh, 2012) mengklasifikasikan blogger menjadi dua grup yaitu: *professional blogger* (PB), dan *Seasonal Blogger* (*Temporary*). *Professional Blogger* adalah blogger yang mengadopsi blog sebagai media yang efektif dan tertarik pada bidang penulisan digital dalam interval waktu yang kontinyu. Sedangkan *Seasonal Blogger* merupakan blogger yang ikut-ikutan membuat postingan dalam periode waktu diskrit.

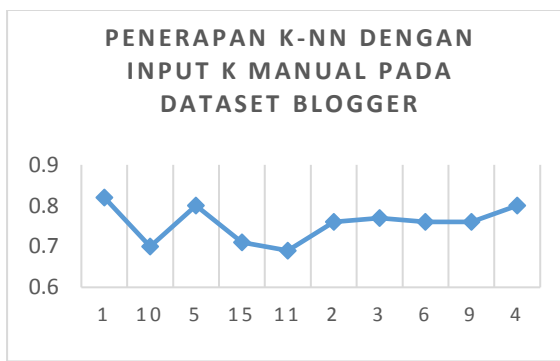
Tahapan kedua, dataset dipecah terlebih dahulu dengan menerapkan *10-fold cross validation* untuk membagi data menjadi dua yaitu 90% *training* dan 10% *testing*. *Cross Validation* adalah estimator yang banyak digunakan untuk mengevaluasi kesalahan prediksi (Bergmeir, Costantini, & Benítez, 2014). Data yang tersedia dibagi menjadi k-set ukuran yang sama, sebagai contoh prosedur *10-fold cross validation* membagi dataset menjadi 10 bagian yang sama besar. Setiap kali fase training, setiap set kecil digunakan sebagai data testing sekali atur dan data selebihnya digunakan untuk training. Jumlah total sample yang salah diakumulasi untuk menghitung akurasi akhir.

Tabel 2. Hasil eksperimen dengan pengaturan nilai k manual

k	ccuracy	Recall	Precision	f-measure
1	0,82	0,652	0,837	0,73
10	0,70	0,078	0,750	0,14
5	0,80	0,495	0,875	0,63
15	0,71	0,112	0,800	0,20
11	0,69	0,112	0,571	0,19
2	0,76	0,445	0,750	0,56
3	0,77	0,470	0,765	0,58
6	0,76	0,303	1	0,47
9	0,76	0,328	0,786	0,46

4 0,80 0,362 1 0,53

Tahapan ketiga, data dilatih dan diuji dengan metode k-NN yang mana nilai k telah diatur oleh GA sebelumnya. Selanjutnya accuracy terbaik dari perolehan nilai k oleh GA akan dipergunakan pada model K-NN+GA untuk dibandingkan dengan model K-NN yang nilai k nya di atur secara manual. Selain ini hasil accuracy K-NN+ GA juga akan dibandingkan dengan hasil accuracy pada penelitian yang sebelumnya. Alur metode penelitian dapat dilihat pada Gambar 6.



Gambar 7. Grafik percobaan k-NN dengan input nilai k Manual pada dataset Blogger.

Tahapan selanjutnya, hasil accuracy k-NN, dan k-NN+ GA juga akan dibandingkan dengan metode lainnya seperti Random Forest, ID3, Random Tree, dan Naive Bayes. Perbandingan beberapa metode ditunjukkan pada Tabel 6, dan Gambar 10.

HASIL DAN PEMBAHASAN

Eksperimen dilakukan menggunakan komputer personal Intel Core i3, 4GB RAM, 500GB HDD, sistem operasi Windows 8.1, dan Rapidminer 8.0.

Pada tahapan pertama, eksperimen dilakukan pada metode k-NN dengan mengatur nilai k secara manual sebanyak 10 kali. Tahapan kedua, eksperimen pada metode k-NN dengan mengatur nilai k secara otomatis, yaitu dengan memberikan jangkauan nilai k pada algoritma genetika.

Seperti yang ditunjukkan pada Tabel 3. Berdasarkan hasil eksperimen pada Tabel 2. Nilai Accuracy yang unggul adalah sebesar 82% dengan nilai k=1. perbandingan akurasi antara metode sebelumnya dengan metode k-NN dan GA+k-NN dapat dilihat pada Tabel 4.

Tabel 3. Hasil Eksperimen optimasi nilai k k-NN oleh algoritma genetika

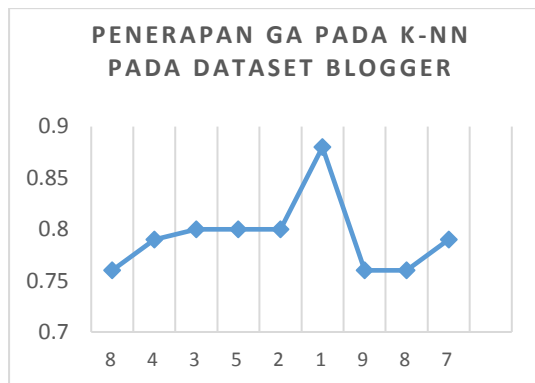
Range	K	Acuracy	Recall	Precision	f-measure
1-100	8	0,76	0,227	1	0,37
1-50	4	0,79	0,357	0,923	0,51
1-30	3	0,80	0,438	0,833	0,57
1-20	5	0,80	0,390	0,875	0,54
1-10	2	0,80	0,423	1	0,59
1-5	1	0,88	0,733	0,857	0,79
5-25	9	0,76	0,323	0,833	0,47
5-20	8	0,76	0,226	1	0,37
5-35	7	0,79	0,405	0,866	0,55

5-45	8	0,76	0,002	0,010	0
------	---	------	-------	-------	---

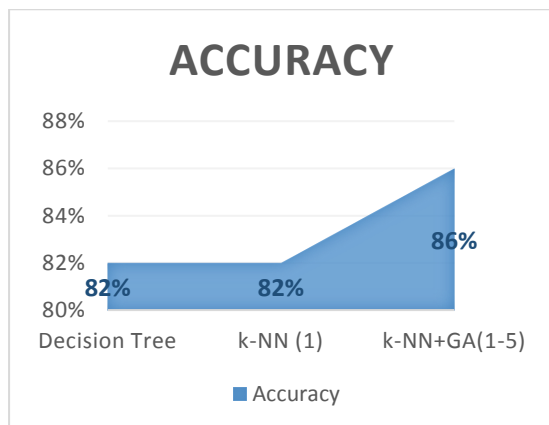
Tabel 4. Perbandingan Nilai Accuracy k-NN dengan k-NN + GA

Machine Learning	Accuracy
Decision Tree (Gharehchopogh, 2012)	82%
k-NN	82%
k-NN + Genetic Algorithm	88%

Jika dibandingkan dengan penelitian sebelumnya. Maka eksperimen k-NN + GA ini berhasil memperbaiki nilai accuracy pada prediksi kecenderungan "Posting Blog". Perbedaan nilai Accuracy dapat ditunjukkan pada Gambar 8. Berdasarkan Tabel 4.



Gambar 8. Grafik percobaan k-NN+GA pada dataset Blogger



Gambar 9. Perbandingan Nilai Akurasi Machine Learning

Berdasarkan hasil eksperimen dengan GA, nilai accuracy paling tinggi adalah 88% dengan range k (1-5) dan hasil nilai k=1.

Tahapan ketiga, peneliti melakukan uji t dua sampel berpasangan dengan membandingkan nilai akurasi pada data uji k-NN yang nilai k nya di setting secara manual, dengan data uji GA+k-NN yang mana nilai k pada k-NN telah dioptimasi oleh GA terlebih dahulu. Uji beda dilakukan dengan menggunakan metode statistik. Untuk menguji hipotesa:

H0: tidak ada perbedaan nilai akurasi antara model k-NN dengan model GA+k-NN

H1: ada perbedaan nilai akurasi antara model k-NN dengan model GA+k-NN

Berdasarkan hasil uji t dua sample berpasangan pada Tabel 5. Nilai t hitung yang diwakili oleh nilai t stat sebesar -1,60578 dan nilai t tabel yang diwakili oleh nilai t critical two tail sebesar 2,26215, maka nilai t hitung < t tabel sehingga H0 diterima dan H1 ditolak.

Sedangkan diketahui nilai probabilitas adalah 0,142785 yang mana nilainya > 0,05 yang artinya tidak terdapat perbedaan yang signifikan antara akurasi k-NN dengan GA+k-NN. Namun kombinasi GA+k-NN tetap memberikan peningkatan akurasi pada BLOGGER dataset.

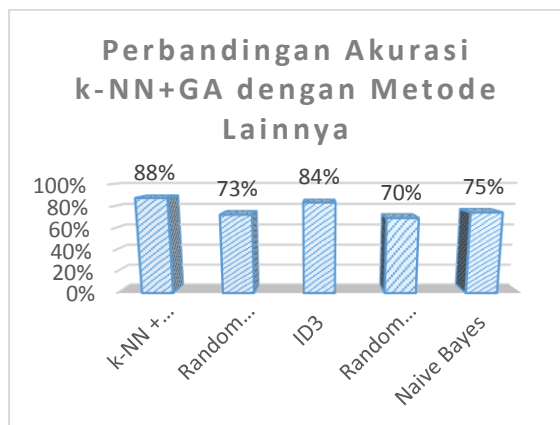
Tabel 5. Hasil Uji Beda Statistik k-Nn dan GA+k-NN

t-Test: Paired Two Sample for Means

	<i>Variable</i> <i>1</i>	<i>Variable</i> <i>2</i>
Mean	0,757	0,79
Variance	0,001979	0,001333
Observations	10	10
Pearson Correlation	-0,28045	
Hypothesized Mean Difference	0	
df	9	
t Stat	-1,60578	
P(T<=t) one-tail	0,071392	
t Critical one-tail	1,833113	
P(T<=t) two-tail	0,142785	
t Critical two-tail	2,262157	

Tabel 6. Perbandingan Nilai Accuracy k-NN + GA dengan Metode Lainnya

Machine Learning	Accuracy
k-NN + Genetic Algorithm	88%
Random Forest	73%
ID3	84%
Random Tree	70%
Naive Bayes	75%



Gambar 10. Grafik perbandingan accuracy k-NN+GA dengan metode lainnya

Berdasarkan Tabel 6. k-NN + GA terbukti lebih unggul dari pada metode lainnya, diikuti ID3 pada posisi kedua, dan Naive Bayes pada posisi ke tiga. Grafik Perbedaan accuracy machine learning ditunjukkan pada Gambar 10.

SIMPULAN DAN SARAN

k-NN dapat mengatasi masalah klasifikasi dengan baik, namun dibalik keunggulan k-NN pengaturan nilai k yang sangat sensitif kerap menjadi kelemahan dari k-NN.

Walaupun tidak ada perbedaan yang signifikan antara accuracy k-NN+GA dengan k-NN namun, Algoritma Genetika

pada penelitian ini telah berhasil memperbaiki akurasi prediksi dengan memberikan hasil nilai k yang optimal bagi k-NN karena gabungan k-NN+GA memiliki nilai akurasi prediksi yang baik jika dibandingkan dengan k-NN tanpa optimasi, juga dengan metode decision tree pada penelitian sebelumnya. Selain itu k-NN + GA juga memiliki accuracy lebih unggul dari pada metode klasifikasi yang lainnya.

Untuk penelitian dimasa mendatang, peneliti percaya bahwa kombinasi k-NN dengan algoritma metaheuristik lainnya dapat meningkatkan akurasi lebih signifikan.

UCAPAN TERIMAKASIH

Penulis mengucapkan terimakasih kepada Suami dan ananda tercinta dan seluruh karyawan BPMA AIPT BSI yang telah memberikan dukungan agar penulis dapat menyelesaikan penelitian ini, Selain itu penulis ingin berterimakasih kepada BSI Group yang memberikan tekanan secara positif untuk menumbuhkan minat penelitian. Ucapan terimakasih juga penulis ucapkan kepada teman-teman pejuang BKD atas semangat-semangat yang selalu diberikan kepada penulis agar dapat mengembangkan penelitian lebih dalam lagi. ucapan terimakasih juga penulis tambatkan kepada teman-teman Group Research Intelegent System yang selalu penuh dengan perbincangan positif mengenai penelitian, sehingga memberikan penulis ide untuk meneliti. Terakhir penulis ingin mengucapkan terimakasih yang sangat mendalam kepada Bapak Romi Satria Wahono yang telah banyak menginspirasi penulis dalam melakukan penelitian di bidang Data Mining.

DAFTAR RUJUKAN

Asim, Y., Shahid, A. R., Malik, A. K., & Raza, B. (2017). Significance of machine learning algorithms in professional blogger ' s classification, 0, 1–13.

Bergmeir, C., Costantini, M., & Benítez, J.

M. (2014). On the usefulness of cross-validation for directional forecast evaluation. *Computational Statistics & Data Analysis*, 76, 132–143.

Bhuvaneswari, P., & Therese, A. B. (2015). Detection of Cancer in Lung With K-NN Classification Using Genetic Algorithm. *Procedia Materials Science*, 10(Cnt 2014), 433–440.

Gharehchopogh, F. S. (2012). Data Mining Application for Cyber Space Users Tendency in Blog Writing : A Case Study, 47(18), 40–46.

Gharehchopogh, F. S., Khaze, S. R., & Maleki, I. (2015). A New Approach in Bloggers Classification with Hybrid of K-Nearest Neighbor and Artificial Neural Network Algorithms, 8(February), 237–246.

Gorunescu, F. (2011). *Intelligent Systems Reference Library*. (Gorunescu, Ed.).

Harafani, H., & Wahono, R. S. (2015). Optimasi Parameter pada Support Vector Machine Berbasis Algoritma Genetika untuk Estimasi Kebakaran Hutan. *Journal of Intelligent Systems*, 1(2).

Harish, N., Mandal, S., Rao, S., & Patil, S. G. (2015). Particle Swarm Optimization based support vector machine for damage level prediction of non-reshaped berm breakwater. *Applied Soft Computing*, 27, 313–321.

Kuncheva, L. I. (1997). FITNESS FUNCTIONS IN EDITING k-NN REFERENCE SET BY GENETIC ALGORITHMS $h \sim M (Z /) h] M (z j)$, 30(6), 1041–1049.

Setiawan, R. (2017). ScienceDirect ScienceDirect Performance Performance Comparison Comparison and and

- Optimization Optimization of of Text
Text Document Document
Classification using Naïve Bayes
Classification Classification using k-
NN and Naïve Bayes Classific.
Procedia Computer Science, 116,
107–112.
- Souza, R., Rittner, L., & Lotufo, R. (2014).
A comparison between k -Optimum
Path Forest and k -Nearest
Neighbors supervised classifiers.
Pattern Recognition Letters, 39, 2–
10.
- Ullah, H., Daud, A., Ishfaq, U., Amjad, T.,
Aljohani, N., Ayyaz, R., & Alowibdi, J.
S. (2017). Computers in Human
Behavior Modelling to identify in fl
uential bloggers in the blogosphere :
A survey. *Computers in Human*
Behavior, 68, 64–82.
- Wakahara, T., & Yamashita, Y. (2014). k -
NN classi fi cation of handwritten
characters via accelerated GAT
correlation, 47, 994–1001.
- Yang, X. (2014). *Nature-Inspired
Optimization Algorithms*. Elsevier.
- Zhang, S., Cheng, D., Deng, Z., Zong, M.,
& Deng, X. (2017). A novel kNN
algorithm with data-driven k
parameter computation, 0, 1–11.
- Zhang, S., Deng, Z., Cheng, D., & Zhu, X.
(2015). Author ' s Accepted
Manuscript Efficient k NN
Classification Algorithm for Big Data
Reference : To appear in :
Neurocomputing. *Neurocomputing*.