

ANALISIS PERBANDINGAN ALGORITMA KLASIFIKASI NNGE, IB3 DAN NNGES

Listian Pratomo¹, Retno Novi Dayawati², Angelina Prima Kurniati³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Nearest neighbor adalah algoritma klasifikasi yang bekerja dengan cara melakukan perbandingan antara suatu kasus baru dengan contoh-contoh sebelumnya. Masalah utama yang dihadapi oleh Nearest neighbor adalah mahal biaya komputasi, hal ini dikarenakan Nearest neighbor menyimpan semua record didalam data latih.

NNGE merupakan perluasan dari Nearest neighbor dengan memperkenalkan konsep exemplar. Tujuan dari penggunaan exemplar adalah untuk mempercepat proses klasifikasi namun tidak mengurangi akurasi. IB3 menggunakan filter untuk menentukan apakah suatu record akan digunakan atau tidak, tujuan dari penggunaan filter adalah untuk mempercepat proses klasifikasi. IB3 memperkenalkan pendekatan statistik untuk menentukan apakah suatu instance akan digunakan pada proses klasifikasi.

NNGEs merupakan kombinasi antara NNGE dan IB3. NNGEs diprediksi akan memiliki akurasi yang jauh lebih baik dari NNGE dan NNGES

Tugas Akhir ini akan melakukan penelitian untuk mengetahui algoritma klasifikasi terbaik dari NNGE, IB3 dan NNGEs. Hasil dari penelitian menunjukkan bahwa NNGEs memiliki akurasi terbaik disetiap dataset. Sedangkan IB3 dan NNGE bekerja baik pada dataset dengan karakteristik tertentu. IB3 memiliki waktu pembangunan model dan klasifikasi tercepat.

Kata Kunci : Nearest Neighbor, NNGE, IB3, NNGEs

Abstract

Abstract Nearest neighbour is a classification algorithm that learns by comparing each new case to previous examples. In the nearest neighbour algorithm, all instances are generally stored in memory during training. When a new query instance is received the memory is searched to find the instance that matches the query instance most closely. The main problem is computationally expensive classifiers since Nearest neighbour save all training instances.

NNGE extends on nearest neighbour by introducing exemplar. The purpose uses exemplar is to reduced classification time without sacrificing akurasi. IB3 applies an akurasi filter on instances and does not use an instance in classifying decisions until it has proved itself to be worthy of being used in the decision making process. IB3 uses filter to reduced classification time.

NNGEs is variant of NNGE that integrate a statistical approach adopted from IB3. NNGEs is predicted will perform better than NNGE and IB3.

This paper deals with the research to know the best classification algorithm from NNGE, IB3 and NNGEs. The outcome of the research shows that the NNGEs indeed have better akurasi in all dataset. NNGE and IB3 have an acceptable performance in specific dataset. IB3 has the fastest time for building model and classifying data.

Keywords : Nearest Neighbor, NNGE, IB3, NNGEs

1. Pendahuluan

1.1 Latar belakang

Klasifikasi merupakan salah satu *task* penting di dalam *data mining*. Salah satu algoritma klasifikasi yang sering digunakan adalah Nearest Neighbour. Keunggulan utama dari Nearest Neighbour adalah mudah untuk dipahami dan mendukung incremental learning. Penambahan *record* pada data latih dapat dilakukan dengan mudah tanpa harus melakukan pembangunan ulang model klasifikasi. Masalah utama yang dihadapi oleh algoritma Nearest Neighbour adalah mahal biaya komputasi. Semakin banyak jumlah instance di dalam data latih maka proses klasifikasi akan semakin lama. NNGE (*Nearest Neighbour with Generalized Exemplar*) dan IB3 merupakan pengembangan dari Nearest Neighbour yang dimaksudkan untuk mengurangi biaya komputasi.

NNGE merupakan pengembangan dari algoritma Nearest Neighbour dengan memperkenalkan konsep general *exemplar*. General *exemplar* merupakan kumpulan dari instance yang memiliki kelas yang sama yang direpresentasikan dengan *rectangle* atau *hyper rectangle* pada data berdimensi banyak. Tujuan penggunaan *exemplar* pada NNGE adalah untuk mempercepat proses klasifikasi namun tidak mengurangi akurasi.

IB3 merupakan anggota dari algoritma IBL (*Instance Based Learning*) yang dikembangkan oleh Aha[1]. IB3 merupakan perbaikan dari IB1 dan IB2. IB3 menggunakan sebuah filter yang berfungsi untuk menyeleksi instance mana saja yang layak digunakan pada proses klasifikasi. Sehingga pada proses klasifikasi data yang digunakan adalah data yang dianggap memiliki kualitas yang baik, sedangkan data yang dianggap tidak layak akan diabaikan. Penggunaan filter pada IB3 juga berakibat jumlah data yang digunakan pada proses klasifikasi menjadi lebih sedikit, sehingga proses klasifikasi dapat berlangsung lebih cepat.

NNGE-S dibangun dengan tujuan untuk meningkatkan performansi dari NNGE. NNGE merupakan suatu classifier yang telah terbukti bekerja dengan baik pada data set yang bersih[9], namun memiliki kinerja yang buruk pada dataset yang memiliki banyak noise, karena proses generalisasi dalam NNGE tidak mengizinkan terdapat konflik antar kelas dalam suatu *rectangle*, jika terdapat instance yang berupa noise maka NNGE akan menjadikannya suatu *rectangle*, sehingga NNGE akan memaksa untuk melakukan *pruning* pada *rectangle* lain yang sebenarnya bukan merupakan noise. Hal ini yang mengakibatkan performansi dari NNGE jauh menurun jika dibandingkan dengan menggunakan data yang tidak mengandung noise. Sedangkan IB3 lebih dapat mentolerir noise, hal ini dikarenakan filter yang digunakan oleh IB3 akan menyeleksi instance mana saja yang akan digunakan untuk proses klasifikasi. Kelebihan dari algoritma IB3 tersebut diharapkan dapat menutupi kelemahan dari NNGE, sehingga NNGE-S mempunyai performansi yang jauh lebih baik.

1.2 Perumusan masalah

Perumusan masalah dalam tugas akhir ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan algoritma NNGE menjadi sebuah classifier.
2. Bagaimana mengimplementasikan algoritma IB3 menjadi sebuah classifier.
3. Bagaimana mengaplikasikan IB3 kedalam NNGE, sehingga menjadi NNGE-S.
3. Bagaimana performansi NNGE-S jika dibandingkan dengan IB3 dan NNGE?

Adapun batasan tugas akhir ini adalah sebagai berikut:

1. Tidak menangani pre-prosesing data.
2. Dataset yang akan digunakan adalah dataset sintetik yang dihasilkan oleh data generator ataupun dataset yang berasal dari UCI Machine Learning Repository.
3. Parameter yang digunakan adalah akurasi dan kecepatan. Hal ini dikarenakan algoritma yang digunakan dikembangkan untuk memperbaiki performansi(waktu dan kecepatan) dari algoritma Nearest Neighbor.

1.3 Tujuan

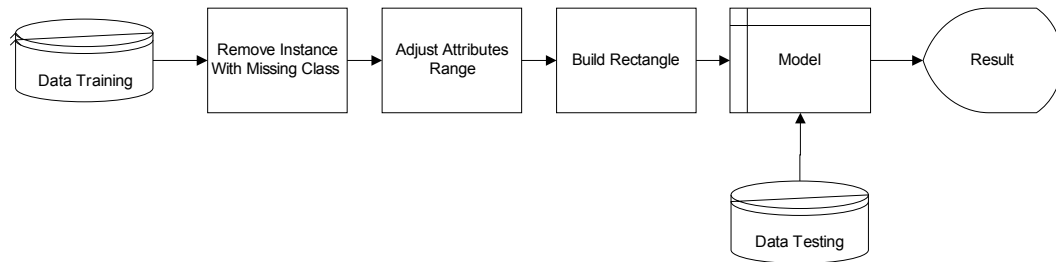
Berdasarkan rumusan masalah di atas, maka tujuan dari tugas akhir ini adalah:

1. Mengetahui kelebihan dan kekurangan masing-masing algoritma.
2. Membangun perangkat lunak yang mengimplementasikan algoritma NNGE, IB3 dan NNGE-S.
3. Menganalisis perbandingan algoritma NNGE, IB3 dan NNGE-S dengan parameter performansi yang berupa kecepatan dan akurasi.

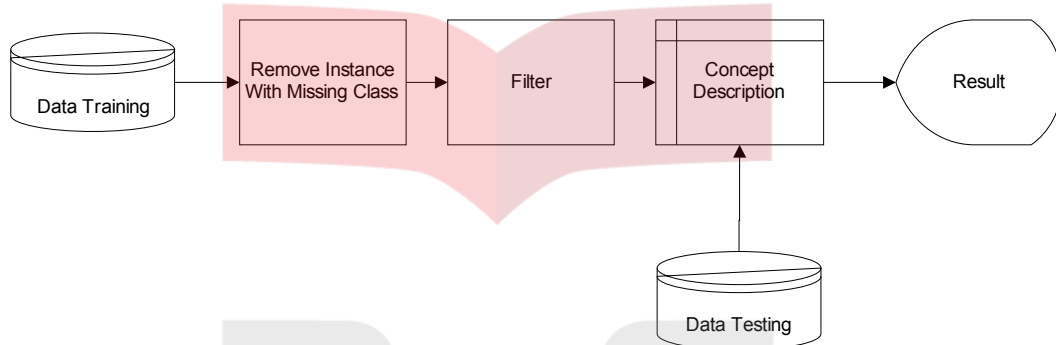
1.4 Metodologi penyelesaian masalah

Metode yang digunakan untuk menyelesaikan tugas akhir ini adalah sebagai berikut:

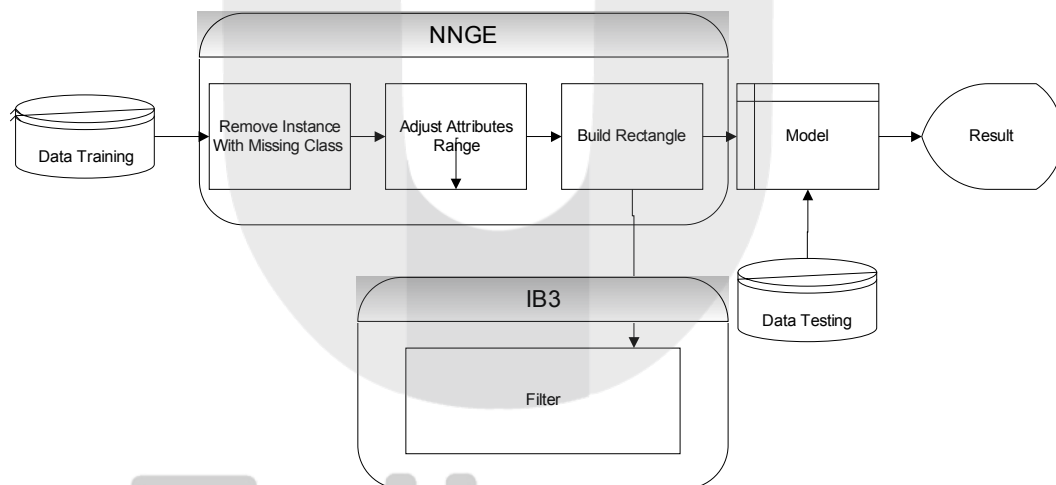
1. Studi literatur dan pengumpulan data
 - a) Pencarian referensi
Mencari referensi dan sumber-sumber lain yang berhubungan dengan algoritma klasifikasi NNGE, IB3 dan NNGE-S.
 - b) Pendalaman materi
Mempelajari dan memahami materi yang berhubungan dengan tugas akhir ini.
 - c) Pengumpulan data
Mengumpulkan *dataset* yang akan digunakan pada tahap analisis.
2. Pembangunan perangkat lunak yang meliputi:
 - a) Analisis dan Perancangan
Melakukan analisis dan perancangan perangkat lunak dengan menggunakan metode berorientasi objek. Input sistem berupa data latih dan data uji. Data latih digunakan untuk pembangunan model klasifikasi, sedangkan data uji digunakan untuk menguji keakuratan model yang telah terbentuk. Output dari system adalah akurasi dari model yang terbentuk serta waktu yang diperlukan.



Gambar 1-2.1 Gambaran sistem untuk algoritma NNGE



Gambar 1-2.2 Gambaran sistem untuk algoritma IB3



Gambar 1-2.3 Gambaran sistem untuk algoritma NNGE-S

b) Pengkodean

Mengimplementasikan perancangan menjadi perangkat lunak dengan menggunakan tool pemrograman.

c) Pengujian

Melakukan pengujian terhadap aplikasi yang akan dibangun, dilakukan dengan pengujian sistem secara fungsional yaitu per proses.

3. Analisis hasil

Melakukan analisis terhadap performansi dari algoritma NNGE, IB3 dan NNGE-S pada berbagai jenis karakteristik data.

4. Pembuatan laporan

Pembuatan laporan Tugas Akhir yang mendokumentasikan tahap-tahap kegiatan dan hasil dalam Tugas Akhir ini.



5 Kesimpulan dan Saran

5.1 Kesimpulan

1. NNGEs memiliki akurasi terbaik untuk setiap dataset yang digunakan, namun memiliki waktu pembentukan model yang paling lama.
2. IB3 memiliki waktu pembangunan model dan klasifikasi paling cepat dibandingkan dengan kedua metode lainnya, akan tetapi akurasi IB3 kurang baik pada saat dataset bersih dari noise.
3. NNGE memiliki waktu nklasifikasi yang paling lama dan memiliki akurasi yang rendah pada datatset dengan noise. Kelebihan dari NNGE adalah tidak perlu menentukan *level confidence* yang akan digunakan.
4. NNGEs dan IB3 memerlukan inputan level confidence untuk proses klasifikasi. Level confidence diperoleh dari percobaan berulang-ulang untuk memperoleh akurasi terbaik.
5. Jumlah noise dapat mempengaruhi kecepatan pembangunan model dan klasifikasi. Namun tidak terdapat pola yang pasti dari pengaruh jumlah noise terhadap kecepatan.

5.2 Saran

1. Perlu dilakukan penelitian lebih lanjut menggunakan parameter lain, seperti besarnya storage yang dipakai oleh masing-masing algoritma. Hal ini dimaksudkan untuk lebih mengetahui kelebihan masing-masing algoritma.
2. Perlu dilakukan penelitian lebih lanjut untuk menggabungkan NNGE dan IB3 dengan metode lain, dengan harapan untuk memperoleh performansi yang jauh lebih baik pada berbagai karakteristik data.

Referensi

- [1] Aha, D. W., Kibler, D and Albert, M. K. Instance-Based Learning Algorithms.
] Machine Learning, 6(1):37-66, 1991.
- [2] Grudic, Greg. 2005. *Nearest Neighbour Learning*.
]
- [3] Han, J., Kamber, M., 2001, *Data Mining: Concepts and Techniques*, USA:
] Morgan Kaufmann, Academic Press.
- [4] Leon, florin. Generalized Nearest Neighbor Method for the Classification of
] Composite Concepts
<http://www.ace.tuiasi.ro/~fleon/papers/35%20NNGEP%20Composite%20Cluj.pdf>
- [5] Martin, B. 1995. Instance-Based Learning: Nearest Neighbour wit Generalisation.
] Master's thesis, University of Waikato.
- [6] Pang-Ning Tan, M. Steinbach and V. Kumar 2006. *Intoduction to Data Mining*
]
- [7] Salzberg, S. 1991. *A Nearest Hyperrectangle Learning Method*. Machine
] Learning, 6(3):251–276
- [8] Weiss, G. M. and Hirsh, H. 1998. *The Problem with Noise and Small Disjuncts*. In
] *ICML*, pages 574–578.
- [9] Wong, Alexander .U.J, Brent Martin, 2005. *Investigating Noise Tolerance in*
] *Generalised Nearest Neighbour Learning*.
- [1] Wettshereck, D. and Dietterich, T. G. *An Experimental Comparison of the*
0] *Nearest-Neighbour and Nearest-HypeAlgorithms*. *Machine Learning*, 19(1):5-27,
1995