

IMPLEMENTASI PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR DAN CLASSIFICATION AND REGRESSION TREE DALAM KLASIFIKASI EVALUASI KINERJA PEGAWAI DI PERUSAHAAN

Afif Ramadhan¹, Wijanarto²

^{1,2}Jurusan Ilmu Komputer, Fakultas Ilmu Komputer Universitas Dian Nuswantoro, Semarang
E-mail: Afiframadhan2@gmail.com¹, Wijanarto.udinus@gmail.com²

Abstrak

Pegawai merupakan salah satu aspek penting dalam keberhasilan berjalannya produksi suatu perusahaan. Dalam suatu perusahaan dibutuhkan sumber daya manusia yang berkualitas untuk meningkatkan kinerja. Dengan kinerja pegawai yang baik maka akan menghasilkan produksi perusahaan secara maksimal sebaliknya jika pegawai tidak menunjukkan kinerja yang baik akan mengurangi angka produksi suatu perusahaan. Kinerja pegawai merupakan nilai kerja ataupun pencapaian kerja pegawai pada periode waktu tertentu dalam melaksanakan pekerjaannya sesuai dengan tanggung jawab dan target produksi. Dari hal tersebut perlu adanya evaluasi yang dilakukan dalam jangka waktu tertentu. Proses evaluasi sangat penting bagi setiap pegawai yang harus dilakukan secara adil, jujur, transparan dan akurat yang sesuai dengan pekerjaan yang dilakukan untuk menentukan kebijakan yang dilakukan perusahaan selanjutnya yang diberikan kepadanya. Dengan memanfaatkan data mining penulis membandingkan algoritma KNN dan CART untuk mengetahui hasil klasifikasi serta akurasi yang lebih baik dari kedua algoritma tersebut ketika memprediksi data evaluasi pegawai di perusahaan. Hasil penelitian menunjukkan bahwa algoritma KNN memiliki akurasi sebesar 85,895% dan algoritma CART memiliki akurasi sebesar 88,46%.

Kata kunci : evaluasi, Algoritma KNN, Algoritma CART, data

Abstract

Employees are one important aspect in the success of a company's production goes. In a company needs qualified human resources to improve the performance. With the good performance of employees will result in the production factory optimally otherwise if an employee does not show a good performance would reduce the rate of production of a company. The performance of employees is the value of the work or the achievement of an employee working for a specific period of time in carrying out the work in accordance with the responsibilities and production targets. From this need for evaluation is done in a certain period. The evaluation process is very important for every employee to be conducted in a fair, honest, transparent and accurate in accordance with the work being done to determine the policy of the company subsequently given to him. By utilizing data mining algorithms authors compared the KNN and CART to determine the classification results and better accuracy of the algorithms are when predicting the evaluation data of employees in the company. The results showed that the KNN algorithm has an accuracy of 85.895% and a CART algorithm has an accuracy of 88.46%.

Keywords: evaluation, KNN algorithm, CART algorithm, data

1. PENDAHULUAN

Pegawai merupakan salah satu aspek penting dalam keberhasilan berjalannya

produksi suatu perusahaan. Dalam suatu perusahaan dibutuhkan sumber daya manusia yang berkualitas untuk meningkatkan kinerja. Dengan kinerja

pegawai yang baik maka akan menghasilkan produksi perusahaan secara maksimal sebaliknya jika pegawai tidak menunjukkan kinerja yang baik akan mengurangi angka produksi suatu perusahaan. Kinerja pegawai merupakan nilai kerja ataupun pencapaian kerja pegawai pada periode waktu tertentu dalam melaksanakan pekerjaannya sesuai dengan tanggung jawab dan target produksi. Dari hal tersebut perlu adanya evaluasi yang dilakukan dalam jangka waktu tertentu. Proses evaluasi sangat penting bagi setiap pegawai yang harus dilakukan secara adil, jujur, transparan dan akurat yang sesuai dengan pekerjaan yang dilakukan untuk menentukan kebijakan yang dilakukan perusahaan selanjutnya yang diberikan kepadanya. Suatu kinerja pegawai dapat ditingkatkan apabila terjadi kesesuaian antara pekerjaan dan kemampuan yang telah dimiliki sehingga dapat mempengaruhi kepuasan kerja. Kepuasan kerja adalah kepuasan dari setiap pegawai terhadap pekerjaan yang dijalannya dimana itu berfungsi sebagai hasil penilaian dari seberapa jauh pekerjaan yang dilakukan secara keseluruhan.

Director Talent dan Rewards Tower Watson Indonesia, Awaldi mengatakan, kemampuan merekrut dan mempertahankan pegawai terbaik terus menjadi tantangan bagi perusahaan Indonesia, antara lain karena mengalami kesulitan dalam memahami faktor-faktor yang mendorong loyalitas dan *engagement* para pekerja profesional di Tanah Air terhadap perusahaan tempat dia bekerja. Faktor yang menyebabkan pegawai Indonesia bertahan dalam sebuah perusahaan selain faktor gaji adalah karena kesempatan pengembangan karir, lingkungan kerja yang nyaman dan lamanya dalam perjalanan ke kantor. Banyak

perusahaan Indonesia memahami ini, sehingga 66 persen pegawai Indonesia cenderung meninggalkan perusahaan tempatnya bekerja dalam kurun waktu 2 tahun, sementara hanya 34 persen mengemukakan niat untuk bertahan di perusahaan tempatnya berkerja saat ini. Indeks komitmen pegawai (comitment index) Indonesia, menurut perhitungan Watson Wyatt(www.portalhr.com, 2010) hanya 57% , lebih rendah 7 poin dibandingkan Asia Pasifik. Itu berarti, tingkat loyalitas kerja pegawai Indonesia termasuk paling rendah di kawasan ini. Seperti yang dikemukakan [1] bahwa loyalitas pegawai Indonesia paling rendah di Asia.

2. METODE PENELITIAN

2.1 Bahan

Dalam penelitian ini dataset yang digunakan adalah data pegawai di suatu perusahaan tekstil pada tahun 2015. Dimana nantinya akan dipilih atribut-atribut yang dibutuhkan untuk proses perhitungan perbandingan.

2.2 Peralatan

Peralatan dalam penelitian ini terdiri dari software dan hardware yang meliputi :

- Software :
 1. Microsoft Exel 2007 .
 2. Widows 7
 3. Notepad++
 4. My SQL
 5. Xampp
- Hardware :
 1. Processor Intel (R) core(TM)i3-2370 M
 2. Memori RAM 2 GB
 3. Hardisk 450 GB

2. 3 Prosedur Pengambilan Data

Pada penelitian ini langkah-langkah pengambilan dataset yaitu dengan menyerahkan surat ijin penelitian yang ditunjukkan kepada PT Delta Merlin, Pondok Sukoharjo. Data yang diperoleh berupa data softcopy karyawan tahun 2015 dengan atribut Nik, nama, bagian, jabatan, unit, regu, absensi, lembur, target produksi.

2. 4 Cara Pengolahan Data

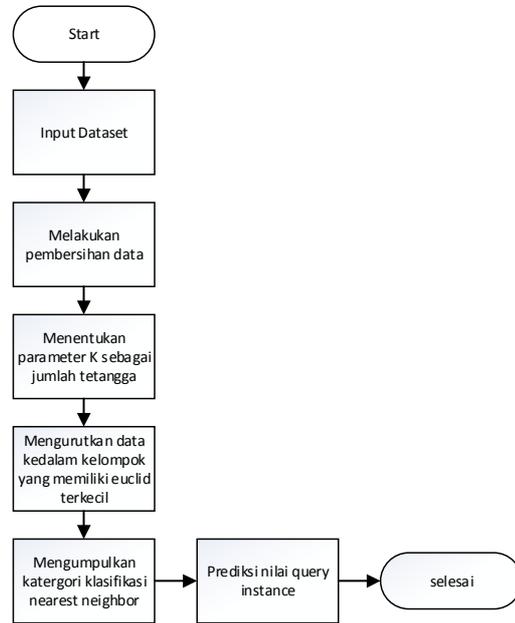
Teknik yang dilakukan dalam proses tahap awal adalah persiapan data yang diambil secara sebanyak 696 data yang akan dijadikan data sampel. Selanjutnya akan dilakukan data cleaning guna menghilangkan data yang tidak konsisten dan data kosong. Apabila terdapat atribut yang tidak digunakan dalam proses klasifikasi maka atributnya dihilangkan sehingga tidak menghambat pada langkah selanjutnya.

2. 5 Metode yang diusulkan

Penelitian ini akan menggunakan dua metode yaitu algoritma K-NN dan CART [2][3][4] . Yang kemudian akan dilakukan perbandingan antara kedua algoritma tersebut. Dimana K-NN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data tetangga paling dekat dengan objek, sedangkan CART merupakan salah satu metode dari salah satu teknik pohon keputusan [5][6]. Selanjutnya kedua metode tersebut akan dibandingkan untuk mengetahui tingkat akurasi paling baik dari kedua algoritma tersebut dalam pengklasifikasian evaluasi kinerja pegawai di perusahaan.

1. Algoritma K-Nearest Neighbor
Langkah-langkah algoritma K-Nearest Neighbor. Berikut adalah Flowchart proses

perhitungan Algoritma K-Nearest Neighbor.



Gambar 1. Langkah Algoritma K-NN

1. Langkah-langkah dalam algoritma K-NN adalah sebagai berikut [7][8]:

- a. Menentukan Parameter K(Jumlah tetangga paling dekat).
- b. Menghitung kuadrat jarak Euclid (query instance) masing-masing objek sampel yang diberikan dengan rumus [9][10] :
- c. $D(x,y)=$

$$\sqrt{\sum_{i=1}^n (x_1 - y_1)^2 + (x_2 - y_2)^2 \dots + (x_n - y_n)^2}$$

- d. Keterangan :
 $D(x,y)$ = Jarak euclidean
 x = data training
 y = data testing
 n = jumlah atribut individu antara 1 s/d n
- e. Mengurutkan objek-objek tersebut kedalam kelompok

yang mempunyai jarak euclid terkecil.

- f. Mengumpulkan kategori Y (klasifikasi nearest Neighbor)
- g. Dengan menggunakan kategori Nearest Neighbor yang paling mayoritas maka dapat diprediksi nilai queri instance yang telah dihitung.

2. Langkah-Langkah Algoritma CART [11][12]

- a. Pembentukan Pohon Klasifikasi
 - Pemilihan (classifier)
 - Penentuan Simpul Terminal
 - Penandaan Label Kelas
- b. Pemangkasan Pohon Klasifikasi
- c. Penentuan Pohon Klasifikasi Optimal



Gambar 2. Langkah Algoritma CART

3. HASIL DAN PEMBAHASAN

Bagian ini berisi uraian hasil penelitian yang telah dilakukan, analisis hasil penelitian dan pembahasan hasil penelitian.

3.1 Persiapan Data

Tahapan awal yang dilakukan dalam penelitian ini adalah persiapan data yang diperoleh dari PT Delta Merlin. Data yang diperoleh merupakan HRD Produksi tahun 2015, data pegawai perusahaan berjumlah 696 data.

3.2 Cleaning Data

Dari hasil data yang diperoleh diperoleh sejumlah permasalahan data terutama data yang tidak lengkap. Untuk itu maka dilakukan proses cleaning data untuk menghilangkan data yang bermasalah. Berikut tabel 1 merupakan data pegawai yang dipakai untuk menentukan bonus pegawai

Pada proses cleaning ini dilakukan penghapusan record dan variabel karena pada pemrosesan data mining ada variabel yang tidak berpengaruh dan tidak digunakan seperti atribut Nama, Bagian, Jabatan, Unit dan Regu. Seperti yang terdapat pada metode penelitian atribut-atribut tersebut tidak digunakan dalam prediksi penentuan bonus pegawai, sehingga tidak mengganggu pada saat proses nantinya. Dari proses cleaning didapatkan data sebesar 693 data yang akan digunakan dalam proses klasifikasi.

Dalam perhitungan KNN ada beberapa langkah yang harus dilakukan, berikut adalah langkah-langkahnya :

1. Menentukan parameter K (jumlah tetangga terdekat). Dalam uji coba pertama ini kita gunakan K=5. Penulis memilih K=5 (ganjil) agar tidak ditemui hasil distance dengan jumlah yang sama sehingga dapat

mengganggu dalam proses klasifikasi.

2. Hitung *square distance to query distance* dengan menghitung jarak pada masing-masing atribut dengan mencari selisih antara atribut data training dan data testing.

$$D(x,y) = \sqrt{\sum_{i=1}^n (x_1 - y_1)^2 + (x_2 - y_2)^2 \dots + (x_n - y_n)^2}$$

Keterangan :

D(x,y)= Jarak euclidean

x = data training

y = data testing

n = jumlah atribut individu antara 1 s/d n

Dari hasil perhitungan rumus mencari jarak diatas didapatkan hasil seperti pada tabel dibawah ini. Dimana hasil tersebut belum dilakukan pengurutan sesuai dengan jarak terkecil.

Tabel 1: Hasil Perhitungan Jarak

NIK	Keterangan	Jarak
646	Dapat	3.0
2937	Dapat	1.0
5515	Dapat	1.0
900	Dapat	1.0
3068	Dapat	1.0
1479	Dapat	1.0
757	Dapat	1.0
1660	Dapat	1.0
.....
6004	Dapat	1.0

3. Lakukan klasifikasi berdasarkan nilai parameter K yaitu K=5
Dengan mencari atau menentukan nilai dari jarak terkecil.

Tabel 2: Pengklasifikasian 5 jarak terdekat

NIK	Absensi	Lembur	Target produksi	Keterangan
4209	0	1	90	Dapat

4. Setelah didapatkan 5 jarak terdekat kemudian tentukan apakah data tersebut mendapat keterangan "dapat" atau "tidak dapat" dengan melihat pada kolom keterangan. Karena tabel diatas jumlah keterangan "dapat" > lebih besar daripada keterangan "tidak dapat" maka tabel data test pertama diatas adalah "dapat"

Tabel 3: Hasil Klasifikasi Data Test 1

NIK	Keterangan	Jarak
2937	Dapat	1.0
5515	Dapat	1.0
900	Dapat	1.0
3068	Dapat	1.0
1479	Dapat	1.0

Percobaan tersebut dilakukan hingga seluruh data testing terhitung.

5. Selanjutnya membandingkan antara data hasil klasifikasi dengan data asli.

Tabel 4: Hasil Akurasi Algoritma KNN dengan data tes 105

kelas sebenarnya	total kelas	prestasi akurasi	dapat	tidak dapat
dapat	66	100	66	0
tidak dapat	39	71.79	11	28
total	105			
akurasi keseluruhan		85.895		

3.3 Perhitungan Algoritma CART

Perhitungan menggunakan Algoritma CART berbeda dengan menggunakan Algoritma KNN. Pada Algoritma CART harus menggunakan pohon keputusan

yang kemudian akan menjadi rule untuk memprediksi penentuan bonus pegawai.

Untuk masalah klasifikasi ini terdapat 2 jenis variabel yaitu target variabel (yang menunjukkan sebuah objek masuk dalam kategori mana) dan variabel prediktor (data para objek yang akan dijadikan dasar untuk memperkirakan objek tersebut masuk kelompok mana). Dalam kasus ini penulis menggunakan 693 data training untuk melakukan pembentukan rule atau pohon keputusan untuk penentuan bonus pegawai.

Tabel 5: Tingkat Akurasi CART Dengan Data Testing 105 data

kelas sebenarnya	total kelas	presentasi akurasi	dapat	tidak dapat
dapat	66	100	66	0
tidak dapat	39	76.92	9	30
total				
akurasi keseluruhan		88.46		

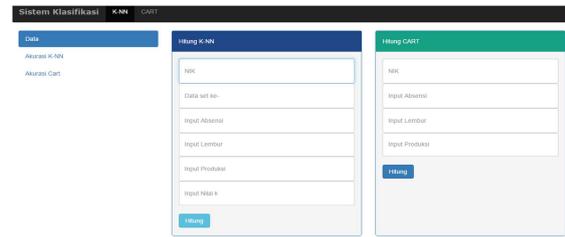
3.4 Hasil Prototype

Untuk membantu dalam memprediksi bonus pegawai, dirancanglah sebuah *prototype*. *Prototype* ini dibuat menggunakan bahasa pemrograman php dan menggunakan database mysql. *Prototype* ini nantinya akan mendapat input dari user berupa nilai K, NIK, absensi, lembur dan target produksi kemudian akan diproses lalu muncul prediksi apakah pegawai tersebut mendapatkan bonus atau tidak.

3.5 Perancangan Tampilan Prototype

Di bawah ini adalah rancangan input pada *prototype* KNN maupun CART. Pada *prototype* ini terdapat tiga buah menu yaitu menu akurasi perbandingan antara KNN dan CART, menu KNN dan menu CART. Pada menu perbandingan akurasi user hanya akan melakukan penekanan tombol hitung perbandingan. Sedangkan untuk menu KNN user akan melakukan inputan berupa nilai K, NIK, absensi, lembur, target produksi. Saat lakukan klik hitung maka akan muncul hasil

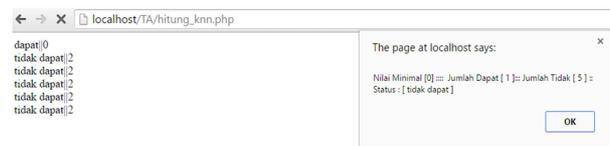
apakah pegawai tersebut mendapatkan bonus atau tidak.



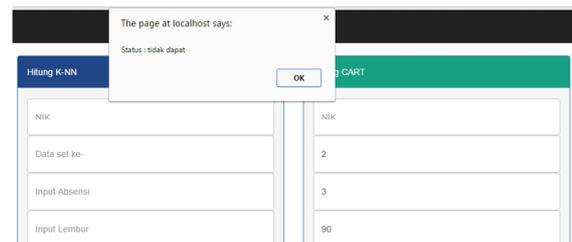
Gambar 3. Halaman Utama Sistem Klasifikasi KNN dan CART

3.6 Implementasi Prototype

Setelah memiliki rancangan *prototype* maka selanjutnya adalah implementasi *prototype*. Pada implementasi bisa dilakukan berdasarkan data training. Pada KNN caranya tinggal memasukan NIK, absensi, lembur, target produksi dan nilai K. Sedangkan pada CART user hanya memasukan NIK, absensi, lembur dan target produksi.



Gambar 4. Implementasi KNN



Gambar 5. Implementasi CART

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan oleh peneliti maka dapat disimpulkan beberapa hal sebagai berikut :

1. Penelitian ini menghasilkan implementasi klasifikasi penerimaan bonus pegawai pada

- perusahaan dengan algoritma KNN dan CART.
2. Dari hasil percobaan prediksi bonus pegawai dengan metode KNN dan CART menggunakan 105 data testing dan 588 data training diperoleh akurasi sebesar 85,895 % dengan nilai $K=5$ untuk algoritma KNN dan untuk algoritma CART sebesar 88.46%. Sehingga antara algoritma KNN dan CART yang lebih tepat digunakan untuk memprediksi penerimaan bonus pegawai adalah algoritma CART.
 3. Dari hasil penelitian diatas dapat disimpulkan bahwa penerimaan bonus pegawai dapat diprediksi dengan menggunakan variabel absensi, lembur, target produksi dan keterangan.

5. SARAN

Untuk meningkatkan kinerja dan dan menyempurnakan penelitian yang telah dibuat, peneliti memberikan saran sebagai berikut :

1. Dalam penelitian ini didapatkan bahwa algoritma CART memperoleh keakuratan lebih baik namun harus menentukan candidate split terlebih dahulu untuk mencari performansi dari setiap cabang.
2. Dalam metode K-NN dalam proses pengklasifikasian memerlukan waktu yang lama karena haru menghitung jarak dari tiap sampel uji.

DAFTAR PUSTAKA

- [1] "Intisari Online," [Online]. Available: <http://intisari-online.com/read/survei-66-persen-karyawan-indonesia-akan-resign-dalam-2-tahun-ke-depan>.
- [2] N. S. dan H. Himawan, "Klasifikasi Bidang Kerja Lulusan Menggunakan Algoritma K-Nearest Neighbor Universitas Dian Nuswantoro," 2014.
- [3] N. Krisnadi, H. Bayu dan B. Prihandono, "Algoritma K-Nearest Neighbor Dalam Klasifikasi data Hasil Produksi Kelapa sawit pada PT.Minamas Kecamatan Parindu," 2013.
- [4] Mardiani, "Penerapan Klasifikasi Dengan Algoritma Cart Untuk Prediksi Kuliah Bagi mahasiswa Baru," 2012.
- [5] F. E. Pratiwi dan I. Zain, "Klasifikasi Pengangguran Terbuka Menggunakan Cart Di Provinsi Sulawesi Utara".
- [6] M. Ridwan, H. Suyono dan M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Baiyes Classifier".
- [7] H. Lidiyana, "Penerapan Algoritma KNN Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," 2013.
- [8] A. Rohman, "Model Algoritma Kn- Nearest Neighbor (K-NN) Untuk Prediksi Kelulusan Mahasiswa".
- [9] M. Lestari, "Penerapan Algoritma Klasifikasi Nearest Neighbor Untuk Mendeteksi Penyakit Jantung," 2014.
- [10] N. Syafitri, "PERBANDINGAN METODE K-NEAREST NEIGHBOR (KNN) dan METODE NEAREST CLUSTER CLASSIFIER (NCC) DALAM PENGKLASIFIKASIAN

- KUALITAS BATIK TULIS,” vol. 2, 2010.
- [11] A. Margasari, “Penerapan Metode CART(Classification And Regresion Tree) Dan Analisis Regresi Logistik Biner Pada Klasifikasi Profil Mahasiswa FMIPA Universitas Brawijaya”.
- [12] A. Hartati, I. Zain dan B. S. Suprih Ulama, “Analisis CART (Classification And Regresion Tree) Pada Faktor-Faktor Yang Memperngaruhi Kepala Rumah Tangga Di Jawa Timur Melakukan Urbanisasi,” 2012.