

KOMPARASI ANALISIS PENERIMAAN KARYAWAN MENGGUNAKAN ALGORITMA C4.5, K-NN DAN NAÏVE BAYES BERBASIS PSO

Ayuni Asistiyasari¹, Yosep Nuryaman²
Universitas Bina Sarana Informatika¹, STMIK Cipta Karya Informatika²
Ayuni.yas@bsi.ac.id

ABSTRAK

Divisi HRD mencoba melakukan penerimaan karyawan sebaik mungkin. Namun walaupun sudah melakukan proses penerimaan dengan baik ternyata ditemukan kejadian dimana tidak sedikit karyawan yang mengundurkan diri sebelum masa kontraknya habis. Untuk mengatasi permasalahan tersebut, maka peneliti melakukan analisis penerimaan karyawan yang ada menggunakan algoritma C4.5, K-NN dan *Naïve Bayes* berbasis PSO dengan maksud agar dapat mengurangi tingkat *turnover* yang ada. Dari hasil penelitian ini menghasilkan hasil akurasi terbaik menggunakan algoritma *Naïve Bayes* berbasis PSO dengan *accuracy* 91,25%, *precision* 88,50% dan *recall* 94,17% serta nilai grafik AUC 0,903 dibandingkan algoritma C4.5 berbasis PSO yang menghasilkan *accuracy* 86,25%, *precision* 80,83% dan *recall* 68,33% serta nilai grafik AUC 0,530 dan algoritma K-NN berbasis PSO yang menghasilkan *accuracy* 82,50%, *precision* 84,33% dan *recall* 75,00% serta nilai grafik AUC 0,796. Simpulan yang diperoleh algoritma *Naïve Bayes* dinyatakan sebagai algoritma terbaik yang dapat digunakan dalam klasifikasi penerimaan karyawan.

Kata kunci: C4.5, KNN, NAIVE BAYES, PSO, Penerimaan Karyawan

ABSTRACT

The HRD Division is trying to do the best possible recruitment of employees. However, despite having done the process of acceptance well it was found that there were not a few employees who resigned before the contract period expired. To overcome this problem, the authors conducted an analysis of existing employee recruitment using the C4.5 algorithm, K-NN and Naïve Bayes based on PSO with the intention to reduce the existing turnover rate. From the results of this research, the best accuracy results are using PSO-based Naïve Bayes algorithm with 91.25% accuracy, precision 88.50% and 94.17% recall and AUC graph value of 0.903 compared to PS4-based C4.5 algorithm that produces accuracy 86, 25%, precision 80.83% and recall 68.33% and AUC graph values of 0.530 and K-NN algorithm based on PSO which resulted in an accuracy of 82.50%, precision 84.33% and 75.00% recall and AUC graphic values of 0.796 . The conclusions obtained by the Naïve Bayes algorithm are stated as the best algorithm that can be used in the classification of employee recruitment.

Keyword: C4.5, KNN, NAIVE BAYES, PSO, Recruitment

PENDAHULUAN

Salah satu bentuk dukungan dari divisi HRD untuk membantu memajukan perusahaan adalah dengan merekrut karyawan-karyawan dengan kualitas terbaik. Dimana karyawan dengan kualitas terbaik tersebut diharapkan memberikan kontribusi yang tinggi terhadap perusahaan nantinya.

Winarso & Arribe (2017:8) mengungkapkan berdasarkan hasil penelitian yang dilakukan terhadap pengembangan sistem seleksi pegawai dan dosen UMRI berbasis e-Rekrutmen menggunakan metode KNN, maka diambil simpulan bahwa metode KNN mampu menyeleksi calon karyawan dan dosen pada tahap seleksi administrasi dengan cara menghitung tingkat kemiripan antara persyaratan dengan data pelamar.

Asistiyasari & Baidawi (2017:8) mengungkapkan model *decision tree* menggunakan algoritma C4.5 yang dievaluasi dengan *confusion matrix* menghasilkan angka akurasi 71,54% +9,13%, *precision* 80,47% +/ -8,16 dan *recall* 83,78% +/ -9,29% serta dengan angka

kurva ROC 0,721; dapat memberikan rekomendasi kepada PT Prismas Jamintara Jakarta dalam menentukan calon pegawai sesuai kriteria yang dibutuhkan oleh klien.

Walaupun sudah menerapkan seleksi karyawan yang ketat dengan harapan mendapatkan karyawan yang terbaik, ternyata permasalahan yang justru muncul ketika karyawan yang sudah lulus seleksi ternyata mengundurkan diri ketika masa kontraknya belum terpenuhi, bahkan terdapat karyawan yang menghilang pada saat proses training. Hal itu tentunya mengakibatkan tingginya *turn over* perusahaan. Hal ini tidak hanya membuat divisi HRD saja yang kesulitan karena harus kembali mengulangi proses pencarian kandidat baru, namun juga divisi yang membutuhkan menjadi terganggu karena kuota SDM nya belum terpenuhi.

Dengan hanya memperhatikan kriteria-kriteria seperti yang disebutkan oleh Winarso & Arribe (2017:8) dan Asistiyasari & Baidawi (2017:8) tentunya tidak dapat menyelesaikan permasalahan tersebut karena tidak menyentuh permasalahan *turnover* perusahaan, karena kebanyakan kriteria lebih kearah kemampuan dari calon karyawan bukan kriteria bahwa karyawan tersebut nantinya akan bertahan lama atau tidak.

Pendekatan yang dilakukan oleh Kurniawan, Purwana, & Widiyastuti (2016) yang meneliti tentang *turnover* suatu perusahaan mengatakan bahwa "nilai *turnover intention* dipengaruhi oleh *variable job insecurity* dan kepuasan kerja sebesar 55,4% dan sisanya 44,6% dipengaruhi oleh variabel lain yang tidak diteliti". Dengan pendekatan ini perusahaan tentunya harus meningkatkan *jobs insecurity* dan juga kepuasan kerja terhadap karyawannya. Namun hal itu bersifat subyektif karena standar *jobs insecurity* dan juga kepuasan kerja setiap karyawan tentunya berbeda-beda.

METODE

Dalam penelitian ini menggunakan model CRISP-DM (*Cross-Industry Standard Proses for Data Mining*), yang terdiri dari 6 tahap yaitu:

1. Fase Pemahaman Bisnis (*Business/Research Understanding Phase*)

Setiap perusahaan tentunya harus melakukan seleksi karyawan dengan baik. Ternyata pada prosesnya setelah proses rekrutmen dilakukan terdapat masalah diantaranya beberapa karyawan yang keluar pada saat kontrak belum habis bahkan beberapa karyawan menghilang pada saat proses training. Oleh sebab itu dibutuhkan metode untuk mengklasifikasi calon karyawan agar tidak membuat *turnover* semakin tinggi.

2. Fase Pemahaman Data (*Data Understanding Phase*)

Untuk pemilihan data diambil dari data karyawan yang bertahan lebih dari 12 bulan, daftar karyawan mengundurkan diri (*resign*) yang dibawah 12 bulan dan calon karyawan yang di tolak. Atribut-atribut yang diambil berasal dari data biodata pribadi karyawan, data penilaian saat wawancara dan tes IQ, adapun atribut-atribut menjadi terlihat pada tabel 1:

Tabel 1. Atribut

| No | Atribut | Nilai |
|----|--------------------|-----------------------------|
| 1 | Usia | 18 sd 40 tahun |
| 2 | Jenis Kelamin | Laki-laki, Perempuan |
| 3 | Status Pernikahan | Lajang, Menikah, Janda/Duda |
| 17 | Penilaian Standar | A,B,C |
| 18 | Penilaian Spesifik | A,B,C |
| 19 | Test IQ | A,B,C |
| 20 | Label | Ditolak, Mundur, Bertahan |

Sumber : Pengolahan Data (2018)

3. Fase Pengolahan Data (*Data Preparation Phase*)

Kemudian dilakukan pembentukan label diterima dan ditolak. Label diterima diambil dari data karyawan yang bertahan lebih dari 12 bulan sedangkan label ditolak diambil dari karyawan yang mundur dibawah 12 bulan serta karyawan yang tidak lulus seleksi.

Tabel 2. Jumlah Data Per Label

| Label | Jumlah |
|-------------|--------|
| Diterima | 45 |
| Ditolak | 66 |
| Jumlah data | 111 |

Sumber : Pengolahan Data (2018)

Selanjutnya dilakukan proses *cleaning* terhadap data-data yang tidak memiliki nilai pada atribut yang dipakai. Sehingga didapat jumlah data yang dapat digunakan dalam penelitian ini adalah sebanyak 80 data dengan rincian 48 data ditolak dan 32 data diterima.

Tabel 3. Jumlah Data Setelah Proses *Cleaning*

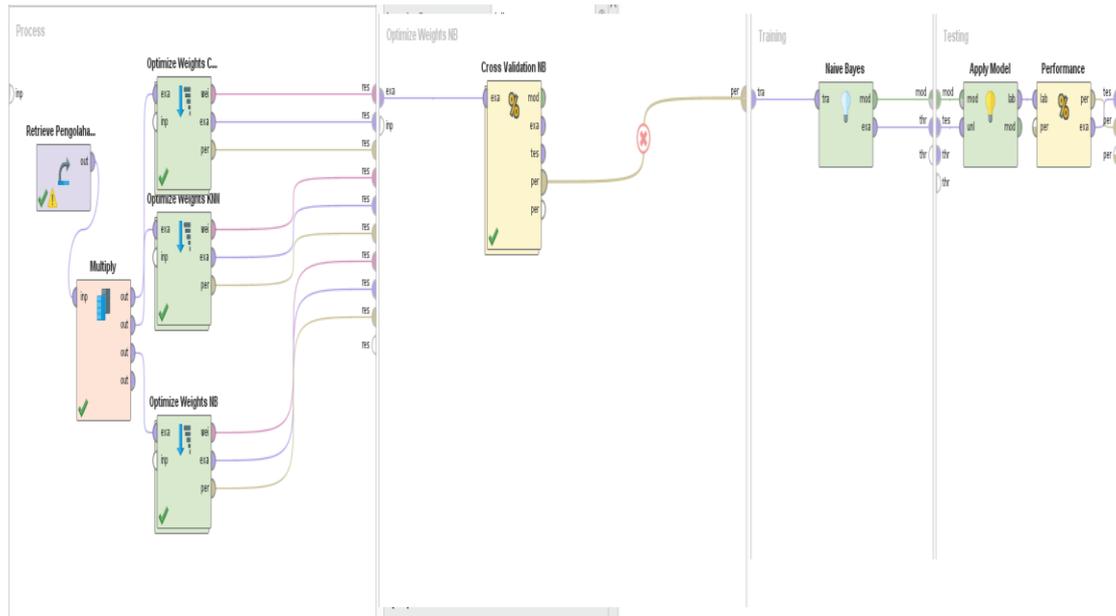
| Label | Jumlah |
|-------------|--------|
| Diterima | 32 |
| Ditolak | 48 |
| Jumlah data | 80 |

Sumber : Pengolahan Data (2018)

Selanjutnya dilakukan pengelompokan pada nilai-nilai atribut, untuk nilai atribut yang bernilai nominal dapat langsung digunakan pada penelitian ini seperti jenis kelamin dengan nilai laki-laki dan perempuan. Namun untuk nilai atribut yang bernilai numerik seperti usia perlu dilakukan pengelompokan.

4. Fase Pemodelan (*Modeling Phase*)

Tahap ini juga disebut tahap *learning* karena pada tahap ini data *training* diklasifikasikan oleh model dan kemudian menghasilkan keputusan. Pada penelitian ini, pembuatan model menggunakan algoritma C4.5, K-NN dan *Naïve Bayes* yang dioptimasi menggunakan PSO seperti yang tampak pada gambar 4.



Gambar 4. Modeling Pada *Rapid Miner*

5. Fase Evaluasi (*Evaluation Phase*)

Pada tahap ini dilakukan pengujian model untuk mendapatkan informasi model yang akurat menggunakan *confusion matrix*.

6. Fase Penyebaran (*Deployment Phase*)

Setelah pembentukan model dan dilakukan analisa dan pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini akan menerapkan hasil penelitian kedalam sebuah aplikasi.

HASIL

Peneliti menggunakan aplikasi *Rapid Miner* untuk melakukan perhitungan dan hasilnya di evaluasi menggunakan *confusion matrix*. Berikut hasil perhitungan *confusion matrix* menggunakan algoritma C4.5 tanpa menggunakan PSO.

accuracy: 80.00% +/- 11.46% (mikro: 80.00%)

| | true Ditolak | true Diterima | class precision |
|----------------|--------------|---------------|-----------------|
| pred. Ditolak | 39 | 7 | 84.78% |
| pred. Diterima | 9 | 25 | 73.53% |
| class recall | 81.25% | 78.12% | |

Gambar 5. *Confusion Matrix* Algoritma C4.5 Tanpa PSO

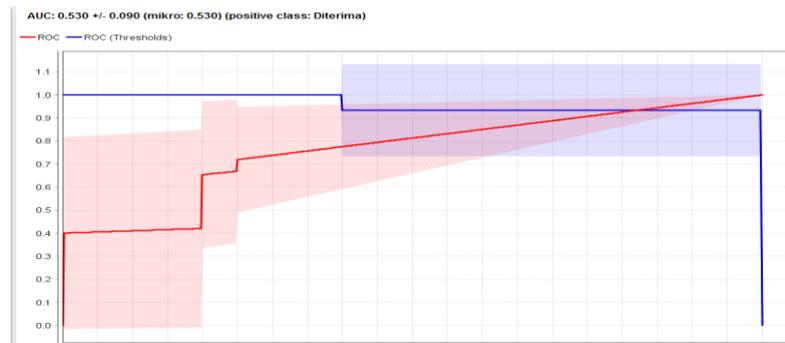
Dan berikut hasil perhitungan *confusion matrix* menggunakan algoritma C4.5 menggunakan PSO.

accuracy: 86.25% +/- 8.75% (mikro: 86.25%)

| | true Ditolak | true Diterima | class precision |
|----------------|--------------|---------------|-----------------|
| pred. Ditolak | 44 | 7 | 86.27% |
| pred. Diterima | 4 | 25 | 86.21% |
| class recall | 91.67% | 78.12% | |

Gambar 6. Confusion Matrix Algoritma C4.5 menggunakan PSO

Nilai AUC yang dihasilkan adalah 0,530 seperti pada gambar 7.



Gambar 7. AUC C4.5 menggunakan PSO

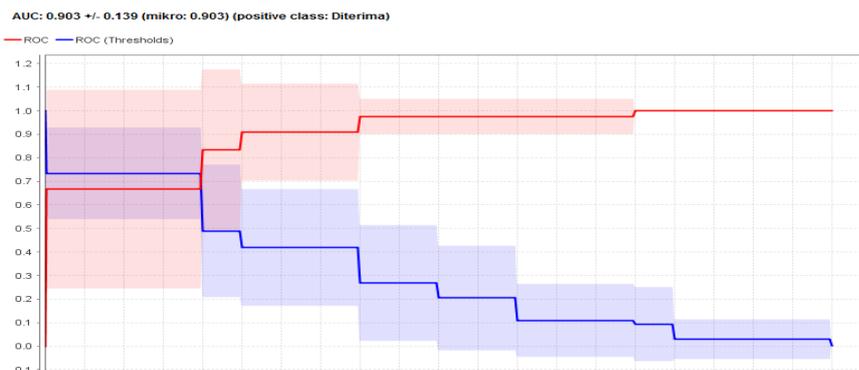
Berikut hasil perhitungan *confusion matrix* menggunakan algoritma *Naïve Bayes* menggunakan PSO.

accuracy: 91.25% +/- 9.76% (mikro: 91.25%)

| | true Ditolak | true Diterima | class precision |
|----------------|--------------|---------------|-----------------|
| pred. Ditolak | 43 | 2 | 95.56% |
| pred. Diterima | 5 | 30 | 85.71% |
| class recall | 89.58% | 93.75% | |

Gambar 8. Confusion Matrix Algoritma Naïve Bayes menggunakan PSO

Dan nilai AUC yang dihasilkan adalah 0,903 seperti pada gambar 9.



Gambar 9. AUC Algoritma *Naïve Bayes* menggunakan PSO

Berikut hasil perhitungan *confusion matrix* menggunakan algoritma K-NN pada K-3 menggunakan PSO.

accuracy: 82.50% +/- 11.46% (mikro: 82.50%)

| | true Ditolak | true Diterima | class precision |
|----------------|--------------|---------------|-----------------|
| pred. Ditolak | 43 | 9 | 82.69% |
| pred. Diterima | 5 | 23 | 82.14% |
| class recall | 89.58% | 71.88% | |

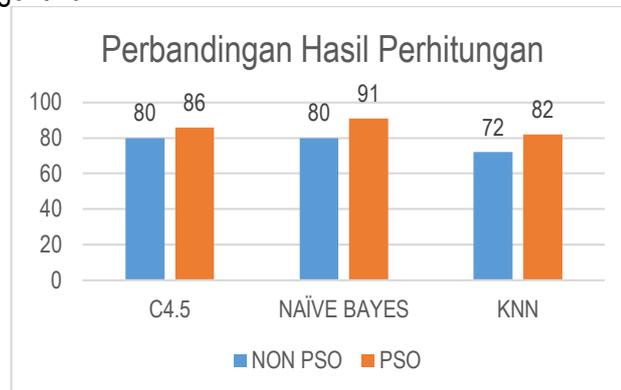
Gambar 10. *Confusion Matrix* Algoritma K-NN menggunakan PSO

Dan nilai AUC yang didapatkan adalah 0,796 seperti pada gambar 11.



Gambar 11. AUC Algoritma K-NN menggunakan PSO

Dalam hasil perhitungan diatas, maka berikut adalah perbandingan hasil perhitungan masing-masing algoritma :



Gambar 12. Perbandingan Hasil Perhitungan

Sesuai gambar 12, dapat dilihat bahwa algoritma *Naïve Bayes* berbasis PSO mampu menghasilkan tingkat akurasi yang lebih baik daripada algoritma C4.5 dan K-NN berbasis PSO.

SIMPULAN

Berdasarkan hasil perhitungan klasifikasi dengan berbagai model yang dioptimasi menggunakan PSO menghasilkan hasil sebagai berikut : 1. Dengan algoritma C4.5 dioptimasi menggunakan PSO yang dievaluasi dengan *confusion matrix* menghasilkan tingkat *accuracy* 86,25%, *precision* 80,83% dan *recall* 68,33% serta nilai grafik AUC 0,530. 2. Dengan algoritma K-NN dioptimasi menggunakan PSO yang dievaluasi dengan *confusion matrix* menghasilkan tingkat *accuracy* 82,50%, *precision* 84,33% dan *recall* 75,00% serta nilai grafik AUC 0,796. 3. Dengan algoritma Naïve Bayes dioptimasi PSO yang dievaluasi dengan *confusion matrix* menghasilkan tingkat *accuracy* 91,25%, *precision* 88,50,% dan *recall* 94,17% serta nilai grafik AUC 0,903.

Maka dapat disimpulkan algoritma terbaik yang dapat digunakan dalam klasifikasi penerimaan karyawan adalah algoritma *Naïve Bayes* dioptimasi PSO yang dievaluasi dengan *confusion matrix* menghasilkan tingkat *accuracy* 91,25%, *precision* 88,50,% dan *recall* 94,17% serta nilai grafik AUC 0,903.

UCAPAN TERIMAKASIH

Ucapan terimakasih kepada perusahaan yang bersedia memberikan izin untuk melakukan penelitian ini, serta kepada Universitas Bina Sarana Informatika dan STMIK Cipta Karya Informatika yang mendukung kami untuk melakukan penelitian ini.

DAFTAR RUJUKAN

- Aprilia, Dennis, Donny Aji Baskoro, Lia Ambarwati dan I Wayan Simri Wicaksana. (2013). Belajar Data Mining dengan Rapid Miner. Jakarta: Gramedia Pustaka Utama.
- Asistiyasari, A., & Baidawi, T. (2017). Analisis Penerimaan Karyawan Posisi Field Collector Menggunakan Algoritma C4 . 5 Pada. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, 2(2), 1–6.
- Dzikrulloh, N. N., & Setiawan, B. D. (2017). Penerapan Metode K–Nearest Neighbor(KNN) dan Metode Weighted Product (WP) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi (Studi Kasus : Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri). *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(5), 378–385. Retrieved from <https://www.safaribooksonline.com/library/view/mastering-machine-learning/9781785283451/ba8bef27-953e-42a4-8180-cea152af8118.xhtml>
- Kusrini dan Emha Taufiq Luthfi. (2009). Algoritma Data Mining. Yogyakarta: Andi Offset.
- Kurniawan, D., Purwana, D. E. ., & Widiyastuti, U. (2016). JOB INSECURITY, KEPUASAN KERJA DAN TURNOVER INTENTION: STUDI PADA KARYAWAN PT “X” DI JAKARTA. *Jurnal Pendidikan Ekonomi Dan Bisnis*, 4(1), 65–78
- Kurniawan, T. C. (2017). Penerapan Algoritma C4.5 Dalam Penerimaan Calon Karyawan PT. Telkom Akses Area Lampung Berbasis Website, 491–501.
- Nafiudin, & Nana Umdiana. (2014). Analisis Faktor-Faktor Yang Mempengaruhi Turnover Intention (Keinginan Berpindah) Karyawan, 8(2), 89–95.
- Winarso, D., & Arribe, E. (2017). Seleksi Pegawai dan Dosen UMRI Berbasis E-Recruitment, 71–80.