

PENERAPAN ALGORITME NEAREST CENTROID NEIGHBOR CLASSIFIER BASED ON K LOCAL MEANS USING HARMONIC MEAN DISTANCE (LMKHNCN) UNTUK KLASIFIKASI HASIL KINERJA PEGAWAI NEGERI SIPIL

Adam Syarif Hidayatullah^{*1}, Fitra A. Bachtiar², Imam Cholissodin³

¹²³Universitas Brawijaya, Malang

Email: ¹adamsyarif219@gmail.ac.id, ² fitra.bachtiar@ub.ac.id, ³imamcs@ub.ac.id

^{*}Penulis Korespondensi

(Naskah masuk: 30 November 2020, diterima untuk diterbitkan: 17 November 2021)

Abstrak

Keberhasilan sebuah perusahaan terjadi karena dapat mengelola sumber daya manusianya dengan baik begitu juga sebaliknya. Salah satu instansi yang mengelola sumber daya manusia menggunakan Manajemen Talenta adalah Badan Kepegawaian Daerah (BKD) kota Malang, dengan mengevaluasi pegawainya setiap tahunnya setelah pekerjaan selesai dilakukan. Hal ini menyebabkan hasil pekerjaan yang telah dilakukan tidak optimal, sehingga perlu identifikasi dini pegawai yang memiliki kinerja dibawah rata – rata sehingga dapat dievaluasi dan meminimalisir hasil pekerjaan yang tidak optimal dengan menggunakan teknik klasifikasi. Penelitian ini menggunakan teknik klasifikasi *Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance* (LMKHNCN). Metode ini merupakan metode modifikasi dari metode *K-Nearest Neighbor* (KNN) dan dibuktikan memiliki performa lebih baik dibandingkan dengan metode aslinya KNN. Dilakukan pengujian *F1-Score* dan akurasi menggunakan *K-Fold Cross Validation* untuk mengetahui persebaran akurasi dan juga pengujian mengenai pengaruh normalisasi karena tidak ada informasi normalisasi pada penelitian sebelumnya. Metode pada kasus ini menghasilkan performa klasifikasi yang baik, dibuktikan bahwa hasil akurasi dan *F1-Score* oleh metode ini berturut – turut ialah mencapai 98,8% dan 98,1%.

Kata kunci: klasifikasi kinerja pegawai, LMKHNCN, data mining

IMPLEMENTATION OF NEAREST CENTROID NEIGHBOR CLASSIFIER BASED ON K LOCAL MEANS USING HARMONIC MEAN DISTANCE (LMKHNCN) FOR EMPLOYEES PERFORMANCE

Abstract

The success of company occurs because is manage human resources well and vice versa. One of institute that mange human resource using Talent Management is Malang city Badan Kepegawaian Daerah (BKD), which evaluates its employee annually after the work is completed. This can cause not optimal work result, so it necessary to early identification of employees who have performance below average performance so that can be evaluated and minimize not optimal result. This study is use classification technique Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance (LMKHNCN). This method is modified base algorithm of K-Nearest Neighbor (KNN). F1-Score and Accuracy using K-Fold Cross Validation to measure performance of this method and normalization testing due to no any information about that in previous study. This method is proven to have better performance compared to it original algorithm KNN. The method in this study has produced good classification performance. The result of classification accuracy and F1-Score by this method reach 98,8% dan 98,1%.

Keywords: employee performance classification, LMKHNCN, data mining

1. PENDAHULUAN

Perusahaan dapat dikatakan baik jika perusahaan tersebut dapat mengelola sumber daya manusianya dengan baik. Perusahaan mengelola sumber daya manusianya agar dapat bertahan di pasar. Salah satu cara mengelola sumber daya

manusia dengan menjaga dan mengelola potensi pegawai yang bisa disebut juga Manajemen Talenta. Manajemen Talenta (Talent Management) merupakan tentang bagaimana mendefinisikan kebutuhan organisasi di masa depan dan tentang bagaimana mengambil tindakan dari kebutuhan kemampuan atau bakat tersebut (Jonason, 2016).

Terbukti pada penelitian (Satria & Gita, 2017) bahwa proses Manajemen Talenta memiliki karyawannya secara intensif menilai kinerja karyawan setiap tahunnya. Hasil dari penilaian tersebut akan diserahkan kepada bagian sub penilaian dan penghargaan kinerja pegawai. Pegawai yang memiliki kinerja yang kurang dari rata – rata kinerja pegawai akan di evaluasi dan di *training* untuk meningkatkan kinerja pegawai tersebut. Tetapi, penilaian akhir kinerja dilakukan setelah pegawai tersebut telah melakukan pekerjaannya setelah itu baru dilakukan proses evaluasi. Hal ini mengakibatkan hasil buah pekerjaan pegawai yang memiliki kinerja dibawah rata – rata tidak optimal. Masalah tersebut dapat diminimalisir jika instansi dapat mengidentifikasi kinerja pegawai secara dini. Salah satu cara identifikasi dini adalah menggunakan teknik klasifikasi dengan memanfaatkan faktor – faktor pengaruh kinerja dan data pada sistem informasi pegawai yang sudah ada.

Klasifikasi merupakan pendekatan teknik secara *supervised learning* atau pendekatan yang mengacu pada dataset yang ada (Aggarwal, 2015). Pemilihan metode klasifikasi dapat mengkaji pada penelitian – penelitian sebelumnya yang berkaitan dengan klasifikasi kinerja. Penelitian yang dilakukan oleh (Dwi, et al., 2018) dan (Raihan & Permana, 2019) berutur – turut mengklasifikan kinerja pegawai universitas menggunakan *neuro fuzzy* dengan akurasi 89% dan klasifikasi kinerja pegawai keamanan menggunakan *K-Nearest Neighbor* dengan akurasi 66,7%. Selanjutnya penelitian yang dilakukan oleh (Setiyorini & Tri, 2018) klasifikasi kinerja siswa mengkomparasi metode *Decision Tree*, *Naïve bayes* dan *K-Nearest Neighbor* (KNN) dengan akurasi tertinggi dimiliki metode KNN. Penelitian klasifikasi kinerja siswa dilanjutkan oleh (Setiyorini & Tri, 2019) dengan menggunakan metode KNN tetapi dengan menambahkan *information gain* dengan hasil kenaikan akurasi sebelum dan setelah penambahan berturut – turut sebesar 74% dan 76%.

Dari beberapa kajian tersebut penelitian ini akan menggunakan metode modifikasi KNN yang harapannya akan menaikkan performa klasifikasi. Metode yang akan dipakai ialah *Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance* (LMKHNCN) yang telah dibuktikan pada penelitian (Mehta, et al., 2018) memiliki nilai performa klasifikasi yang lebih baik dibandingkan metode aslinya KNN dengan tingkat error yang rendah dan stabil jika nilai parameter K yang berubah – ubah.

2. METODE PENELITIAN

2.1. Local Mean K-Nearest Neighbor

Local Mean K-Nearest Neighbor (LMKNN) merupakan metode modifikasi klasifikasi dari *K-Nearest Neighbor* (KNN). LMKNN dapat menurunkan atau mereduksi efek dari outlier yang

ada. Perubahan pada metode LMKNN ada pada parameter k , dimana parameter k pada LMKNN merupakan k jumlah tetangga terdekat pada setiap kelas sedangkan nilai k pada KNN adalah jumlah tetangga terdekat untuk seluruh kelas (Syaliman, et al., 2017). Berikut merupakan langkah – langkah metode LMKNN:

1. Menghitung jarak terdekat data uji ke setiap data latih sejumlah k dari setiap kelas menggunakan metode pengukuran jarak *Euclidean*.
2. Menghitung *local mean vector* menggunakan Persamaan 1 dari setiap kelas terdekat.

$$lm_i = \sqrt{\frac{1}{k} \sum_{j=1}^k d(.)} \quad (1)$$

Keterangan:

lm_i : jarak *local mean vector*

k : jumlah tetangga dari setiap kelas

$d(.)$: jarak *Euclidian*

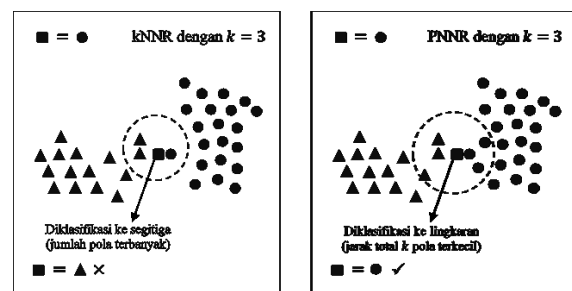
3. Mencari nilai terkecil dari jarak rata- rata m_i menggunakan Persamaan 2 dan mengklasifikasikan x menjadi kelas i .

$$c_i = \text{argmin}(lm_i) \quad (2)$$

Keterangan:

c_i : kelas ke – i

Berikut merupakan gambar ilustrasi dari metode LMKNN dapat dilihat pada Gambar 2.



Gambar 1. KNN (kiri) dan LMKNN (kanan)

2.2. Nearest Centroid Neighbor

Metode *Nearest Centroid Neighbor* (NCN) merupakan metode yang digagas dan diteliti oleh (Chaudhuri, 1996). Pada penelitiannya ia mencetus sebuah metode berdasarkan aturan ketetanggaan dari metode *K-Nearest Neighbor* (KNN), dengan motivasi memperbaiki sensitivitas parameter K dari metode KNN. Sama dengan KNN, NCN juga mencari jarak terdekat k tetangganya dengan data uji (centroid).

Berbeda dengan KNN, NCN membuktikan data uji juga terletak simetris diantara data latih. Sehingga NCN peduli terhadap dua kriteria, yaitu

jarak dan tata letak data. Ilustrasi metode NCN dapat dilihat pada Gambar 2. Berikut ini merupakan langkah – langkah metode NCN (Chaudhuri, 1996):

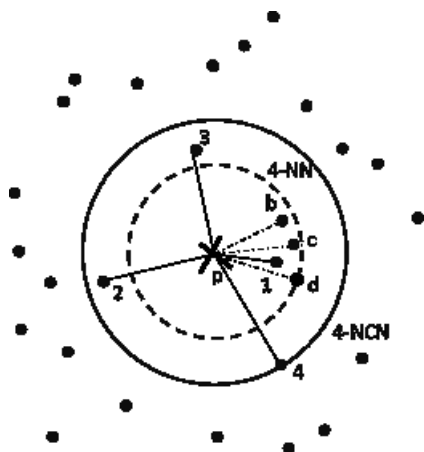
1. Mencari tetangga terdekat pertama pada data uji dengan data latih, menggunakan metode perhitungan. jarak *Euclidian*.
2. Mencari centroid terdekat pada data uji dari tetangga pertama p_1 dengan tetangga ke – k terdekat p_k . Perhitungan *centroid* dapat dilakukan dengan persamaan berikut:

$$c_k = \frac{1}{k} \sum_{j=1}^k p_j \quad (3)$$

Keterangan:

c_k : *centroid* ke – k

p_j : *point* ke – j



Gambar 2. KNN (4-NN) dan NCN (4-NCN)

2.3. Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance

Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance (LMKHNCN) merupakan metode kombinasi antara *Local Mean Nearest Neighbor* (LMKNN) dan *K-Nearest Centroid Neighbor* (KNCN). LMKHNCN mempertimbangkan dua faktor yaitu kemiripan dan jarak yang telah dibahas pada (Mehta, et al., 2018). Menggunakan metode pengukuran jarak menggunakan harmonic distance karena lebih konsisten pada vektor rata – rata *centroid* lokal yang membuat nilai k tidak terlalu sensitif dibandingkan dengan *K-Nearest Neighbor* (KNN) konvensional. Berikut merupakan persamaan metode pengukuran jarak *harmonic distance* yang dapat dilihat pada Persamaan (4).

$$HMDS_i = \frac{k}{\sum_{j=1}^k \frac{1}{d(x, c_k^{NCN})}} \quad (4)$$

Keterangan:

$HMDS_i$: *harmonic mean* ke – i

k : jumlah tetangga terdekat ke k

$d(x, c_k^{NCN})$: jarak *Euclidian*

Tahapan dari metode *Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance* (LMKHNCN) adalah sebagai berikut:

1. Mencari set poin dari *K-Nearest Centroid Neighbor* (KNCN).
2. Menghitung *local mean vector* dari set poin (KNCN) dari setiap kelas.
3. Menghitung jarak *centroid* lokal dari setiap kelas dengan menggunakan *harmonic mean distance* Persamaan 4.
4. Mengklasifikasikan data uji x sebagai kelas c dimana kelas c yang memiliki nilai harmonic mean distance yang paling minimum.

2.4. Pengumpulan Data

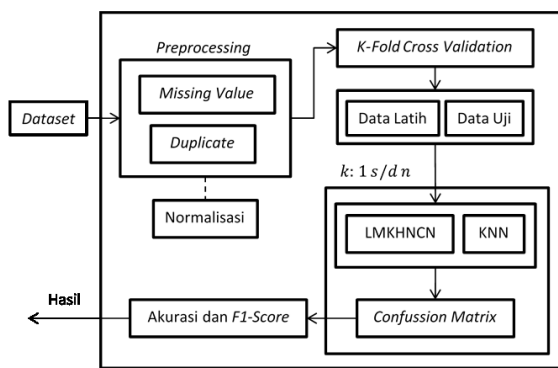
Teknik pengumpulan data pada penelitian penerapan metode *Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance* (LMKHNCN) untuk klasifikasi kinerja pegawai negeri sipil Kota Malang dengan menggunakan data sekunder di dalam sistem informasi pegawai Kota Malang. Pengambilan data dilakukan dengan cara *Web Scrapping* menggunakan *Library Selenium* disetiap data pegawai pada sistem informasi pegawai Kota Malang dengan pengambilan fitur menggunakan penelitian sebelumnya juga pendapat ahli sebanyak 6584 data pegawai. Berikut merupakan fitur dan tipe datanya yang dapat dilihat pada Gambar 3.

Fitur	Tipe Data
Masa Kerja	Numerik
Gaji	Numerik
Jumlah Kenaikan Pangkat	Numerik
Jumlah Penghargaan	Numerik
Orientasi	Numerik
Integritas	Numerik
Disiplin	Numerik
Kerjasama	Numerik
Kawin	Kategorikal / Nominal
Kinerja (Kelas)	Kategorikal / Nominal

Gambar 3. Fitur dataset

2.5. Perancangan

Berikut ini merupakan diagram tahapan – tahapan dari strategi penelitian yang dilakukan dalam klasifikasi pegawai negeri sipil Kota Malang yang dapat dilihat pada Gambar 4.



Gambar 4. Diagram blok perancangan

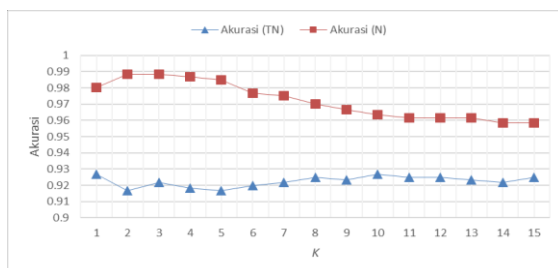
2.6. Teknik Pengujian

Pengujian akurasi dilakukan dengan menggunakan *Confussion Matrix* dan *10-Fold Cross Validation* agar keseluruhan performa model klasifikasi dapat diukur, dimana metode *10-Fold Cross Validation* akan mengevaluasi akurasi dari keseluruhan bagian data. Menguji nilai akurasi dan beberapa percobaan nilai parameter k pada metode *Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance* (LMKHNCN).

3. HASIL DAN PEMBAHASAN

3.1. Pengujian Akurasi dan *F1-Score*

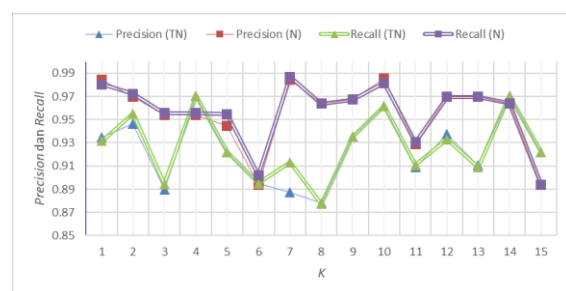
Pada tahap ini akan dilakukan pengujian sensitivitas parameter k dan pengujian perbandingan performa klasifikasi data setelah normalisasi dan tanpa normalisasi pada metode *Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance* (LMKHNCN). Pengujian ini dilakukan agar mengetahui apakah parameter k sensitif jika diubah nilainya dan mengetahui pengaruh performa klasifikasi untuk data menggunakan normalisasi dan tanpa normalisasi. Hasil pengujian akurasi untuk sensitivitas parameter k tanpa normalisasi dapat dilihat pada Gambar 5. N merupakan normalisasi dan TN adalah tanpa normalisasi.



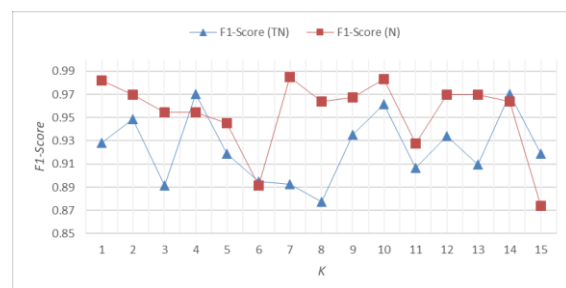
Gambar 5. Grafik Hasil Pengujian Akurasi Sebelum dan Sesudah Normalisasi

Pada Gambar 5 di atas terlihat bahwa akurasi pada data yang dinormalisasi secara keseluruhan memiliki akurasi yang lebih tinggi dibandingkan dengan data yang tidak dinormalisasi. Hal ini dikarenakan normalisasi pada tahapan preprocessing akan menyamakan jarak setiap fiturnya sehingga fitur yang memiliki jarak yang jauh tidak mendominasi yang mengakibatkan kesalahan klasifikasi.

Selanjutnya pengujian performa dengan menggunakan *F1-Score*, pengujian *F1-Score* lebih baik digunakan untuk pengukuran performa dari data yang tidak seimbang (Suyanto, 2017). Hasil dari pengujian *F1-Score* tanpa normalisasi dan menggunakan normalisasi berturut – turut dapat dilihat pada Gambar 6 dan 7.



Gambar 6. Grafik Hasil Pengujian Precision dan Recall

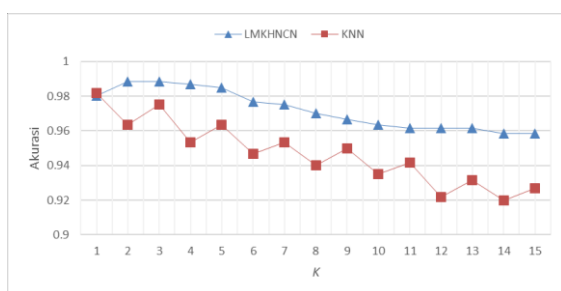


Gambar 7. Grafik Hasil Pengujian F1-Score

Sama dengan kesimpulan akurasi diatas hal ini dikarenakan normalisasi pada tahapan preprocessing akan menyamakan jarak setiap fiturnya sehingga fitur yang memiliki jarak yang jauh tidak mendominasi yang mengakibatkan kesalahan klasifikasi.

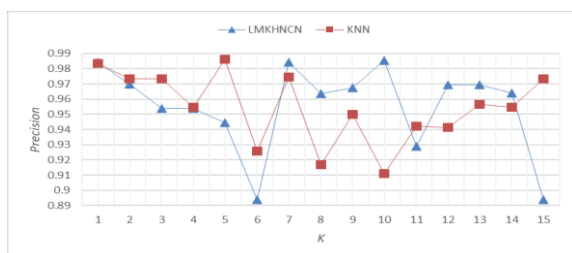
3.2. Pengujian LMKHNCN dengan KNN

Pada tahap ini akan dilakukan pengujian perbandingan metode LMKHNCN dengan metode KNN agar mengetahui apakah metode LMKHNCN lebih baik mengklasifikasikan dibandingkan dengan metode yaitu KNN. Hasil pengujian dapat dilihat pada Gambar 8.

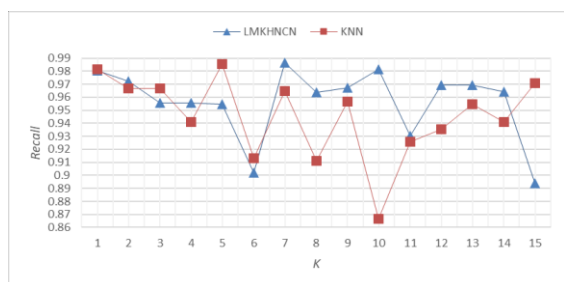


Gambar 8. Grafik Hasil Pengujian Akurasi

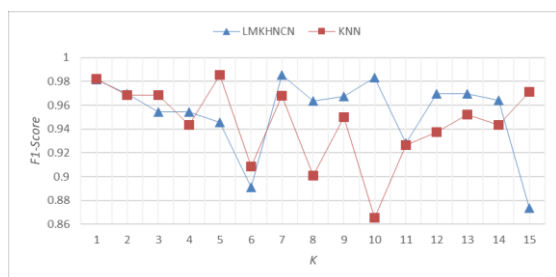
Pada Gambar 8 di atas terlihat bahwa akurasi pada metode LMKHNCN hampir keseluruhan memiliki akurasi yang lebih tinggi dibandingkan dengan metode KNN. Hal ini dikarenakan proses dalam metode LMKHNCN tidak hanya mempertimbangkan kedekatan data juga mempertimbangkan tata letak data sehingga performa dari klasifikasi meningkat dibandingkan dengan metode aslinya KNN. Selanjutnya pengujian performa dengan menggunakan F1-Score dapat dilihat pada Gambar 9, 10, dan 11.



Gambar 9. Grafik Hasil Pengujian Precision



Gambar 10. Grafik Hasil Pengujian Recall



Gambar 11. Grafik Hasil Pengujian F1-Score

Sama halnya dengan kesimpulan nilai akurasi metode LMKHNCN sebagian besar nilai *precision*, *recall* dan *F1-Score* dari setiap nilai *K* memiliki nilai yang lebih tinggi dibandingkan metode KNN. Tetapi, ada beberapa nilai yang diungguli oleh KNN.

Hal ini dikarenakan proses dalam metode LMKHNCN tidak hanya mempertimbangkan kedekatan data juga mempertimbangkan tata letak data sehingga performa dari klasifikasi meningkat dibandingkan dengan metode aslinya KNN.

4. KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian penerapan metode *Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance* (LMKHNCN) pada klasifikasi kinerja pegawai negeri sipil pemerintah kota malang adalah penggunaan normalisasi pada saat preprocessing berpengaruh positif terhadap performa akurasi maupun *F1-Score* khususnya pada metode yang menggunakan perhitungan jarak pada algoritmenya. Hal ini dibuktikan pada Gambar 5 dan 7 performa akurasi dan *F1-Score* pada data yang menggunakan normalisasi secara keseluruhan nilai *K* memiliki performa yang lebih baik dibandingkan dengan data tanpa menggunakan normalisasi. Nilai akurasi dan *F1-Score* tertinggi tanpa normalisasi berturut – turut yakni 92,6% dan 97,05% sedangkan data yang dinormalisasi nilai akurasi dan nilai *F1-Score* tertinggi berturut – turut adalah 98,8% dan 98,1%.

Kestabilan performa akurasi dan *F1-Score* pada model klasifikasi metode *Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance* (LMKHNCN) dengan metode aslinya *K-Nearest Neighbor* (KNN) pada perubahan parameter *K* pada kasus ini keduanya terlihat stabil. Terlihat pada Gambar 8 dan 11 perubahan nilai akurasi dan *F1-Score* pada keduanya tidak berubah secara signifikan setiap perubahan pada parameter *K*. Hal ini terjadi karena data pada kasus ini yang terdistribusi secara merata, sehingga performa model klasifikasi yang tinggi juga stabil.

Perbandingan performa akurasi pada model klasifikasi metode *Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance* (LMKHNCN) dengan metode aslinya *K-Nearest Neighbor* (KNN), metode LMKHNCN secara keseluruhan lebih baik dibandingkan dengan metode aslinya dengan nilai akurasi tertinggi berturut- turut 98,8% dan 98,1%.

DAFTAR PUSTAKA

- AGGARWAL, C., 2015. *Data Mining: The Text Book*. Switzerland: Springer International Publishing.
- CHAUDHURI, 1996. *A New Definition of Neighborhood of a Point in Multidimensional Space*. Pattern Recognition Letters, Volume 17, pp. 11-17.
- DWI, R., BUDI, P. & YUDANINTYAS, E., 2018. Klasifikasi Kinerja Pegawai Universitas X dengan Pendekatan Neuro Fuzzy. Jurnal EECCIS, 12(1).

- JONASON, K., 2016. *Strategic Talent Management*. Denmark: DK-2900
- MEHTA, S., SHEN, X. & GOU, J., 2018. *A New Nearest Centroid Neighbor Classifier Based on K Local Means Using Harmonic Mean Distance*. Information, 9(9), pp. 1-16.
- PUTRI, IR., CHOLISSODIN, I., & SETIAWAN, BD., 2015, 'Optimasi Metode *Adaptive Fuzzy K-Nearest Neighbor* Dengan *Particle Swarm Optimization* Untuk Klasifikasi Status Sosial Ekonomi Keluarga', DORO: Repository Jurnal Mahasiswa PTIIK Universitas Brawijaya, vol. 5, no. 3
- RAIHAN, M. & PERMANA, A., 2019. Penerapan Metode *K-Nearest Neighbor* untuk Klasifikasi Kinerja Satpam Berbasis Web. Jurnal Teknologi dan Manajemen Informatika, 5(1).
- SATRIA, N. & GITA, H., 2017. Pengaruh Manajemen Talenta terhadap Kinerja Karyawan Lintasarta Kota Jakarta. Jurnal Manajemen Indonesia, 17(3).
- SETIYORINI, T. & TRI, R., 2018. Komparasi Metode *Decision Tree*, *Naive Bayes* dan *K-Nearest Neighbor* pada Klasifikasi Kinerja Siswa. Jurnal TECHNO Nusa Mandiri, 15(2).
- SETIYORINI, T. & TRI, R., 2019. Penerapan Metode *K-Nearest Neighbor* dan *Information Gain* pada Klasifikasi Kinerja Siswa. Jurnal Ilmu Pengetahuan dan Teknologi komputer, 5(1).
- SYALIMAN, U., NABABAN, E. & SITOMPUL, O., 2017. *Improving The Accuracy Of K-Nearest Neighbor Using Local Mean Based And Distance Weight*. Journal of Physics, Volume 978, pp. 1-6.
- SUYANTO, 2017. Data Mining: untuk Klasifikasi dan Klasteriasi Data. Bandung: Informatika.
- THEODORIDIS, S. & KOUTROMBAS, K., 2003. *Pattern Recognition Second Edition*. USA: Elsevier.
- ZAKI, M. & MEIRA, W., 2014. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. USA: Cambridge University Press.