

PERBANDINGAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI WAKTU TUNGGU ALUMNI DALAM MEMPEROLEH PEKERJAAN (Study Kasus SMKS PGRI 2 PRINGSEWU)

Muhamad Brilliant¹, Iis Ariska Nurhasanah², Dedeh Rahmadaniah³

^{1,2,3}Institut Teknologi dan Bisnis Diniyyah Lampung

muhamadbrilliant9393@gmail.com

Article History:

Received: Juli 042023

Revised: Agustus 04, 2023

Accepted: Agustus 6, 2023

Published: Agustus 7, 2023

Keywords: *Data mining, Algoritma Naïve Bayes (NB), K-Nearest Neighbors (KNN)*

***Corresponding author:**

muhamadbrilliant9393@gmail.com¹

Abstract: *Klasifikasi waktu tunggu alumni dalam memperoleh pekerjaan merupakan aspek penting dalam evaluasi keberhasilan sebuah lembaga pendidikan. Dalam konteks SMKS PGRI 2 Pringsewu, penelitian ini bertujuan untuk membandingkan performa dua algoritma klasifikasi, yaitu Naïve Bayes dan K-Nearest Neighbor (KNN), dalam memprediksi waktu tunggu alumni dalam memperoleh pekerjaan. Dalam penelitian ini, atribut-atribut seperti Jurusan, Nilai rata-rata, Prestasi Akademik/Non-Akademik dan Ekstrakurikuler digunakan sebagai input untuk memprediksi waktu tunggu alumni dalam memperoleh pekerjaan. Naïve Bayes memberikan prediksi berdasarkan probabilitas kelas dan probabilitas fitur yang terkait dengan kelas tersebut, sementara KNN mencari K alumni dengan atribut mirip untuk memprediksi waktu tunggu berdasarkan waktu tunggu alumni terdekat. Hasil perbandingan ini memberikan informasi penting dalam memilih algoritma yang paling sesuai untuk melakukan klasifikasi waktu tunggu alumni dalam memperoleh pekerjaan. Penelitian ini dapat memberikan wawasan bagi lembaga pendidikan dalam memantau dan meningkatkan tingkat kesuksesan alumni dalam memasuki dunia kerja.*

1. Introduction

Dalam dunia pendidikan, tingkat kesuksesan alumni dalam memperoleh pekerjaan setelah lulus merupakan indikator penting bagi lembaga pendidikan dalam mengevaluasi kualitas pendidikan yang diberikan. Oleh karena itu, lembaga pendidikan seperti SMKS PGRI 2 Pringsewu perlu memahami faktor-faktor yang mempengaruhi waktu tunggu alumni dalam memperoleh pekerjaan.

Klasifikasi waktu tunggu alumni dalam memperoleh pekerjaan adalah suatu teknik yang dapat digunakan untuk memprediksi lamanya waktu yang diperlukan bagi seorang alumni untuk mendapatkan pekerjaan setelah lulus. Dalam konteks ini, dapat digunakan dua algoritma klasifikasi yang populer, yaitu Naïve Bayes dan K-Nearest Neighbor (KNN).

Dalam konteks SMKS PGRI 2 Pringsewu, algoritma Naïve Bayes dapat digunakan untuk memperkirakan waktu

tunggu alumni dalam memperoleh pekerjaan berdasarkan atribut-atribut seperti Jurusan, Nilai rata-rata, Prestasi Akademik/Non-Akademik dan Ekstrakurikuler. Algoritma ini dapat memberikan probabilitas waktu tunggu berdasarkan atribut-atribut tersebut. Di sisi lain, algoritma KNN dapat memprediksi waktu tunggu alumni berdasarkan atribut-atribut yang telah diketahui dari alumni yang sudah lulus sebelumnya.

Dalam rangka memilih algoritma yang paling sesuai untuk melakukan klasifikasi waktu tunggu alumni dalam memperoleh pekerjaan, perlu dilakukan perbandingan performa antara algoritma Naïve Bayes dan KNN. Hal ini akan memberikan wawasan yang penting bagi SMKS PGRI 2 Pringsewu dalam meningkatkan tingkat kesuksesan alumni dalam memasuki dunia kerja.

2. Tinjauan Pustaka

2.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar [1]. Hal penting yang terkait di dalam data mining adalah:

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat[2].

2.2 Algoritma naïve bayes

Klasifikasi *Bayesian* adalah klasifikasi statistik yang bisa diklasifikasi probabilitas sebuah kelas. Klasifikasi Bayesian ini dihitung berdasarkan *Teorema Bayes*. *Teorema Bayes* adalah perhitungan statistik dengan menghitung probabilitas kemiripan kasus lama yang ada dibasis kasus dengan kasus baru. *Teorema Bayes* tingkat

akurasi yang tinggi dan kecepatan yang baik ketika diterapkan pada database yang besar[3].

2.3 Algoritma *K-Nearest Neighbors* (KNN)

Menurut [4] algoritma KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Ketepatan algoritma KNN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi.

3. Metode Penelitian

3.1 Metode Pengumpulan Data

Pada proses penulisan jurnal saat ini menggunakan 3 (tiga) tahapan yang digunakan adalah sebagai berikut:

a. Metode Observasi

Teknik pengumpulan data dengan mengadakan pengamatan dan penelitian secara langsung (Observasi) terhadap obyek yang diteliti melalui wawancara dengan staf sekolahan untuk mendapatkan informasi berdasarkan data yang di butuhkan khususnya data siswa SMKS PGRI 2 Pringsewu. Selain itu untuk merekam berbagai fenomena (situasi dan kondisi) yang terjadi mengenai pendidikan di SMKS PGRI 2 Pringsewu.

b. Metode Dokumentasi

Metode ini juga berfungsi sebagai pengumpulan data-data pendukung seperti gambar-gambar atau dokumen yang lainnya sebagai penjelasan tentang data-data yang sudah diperoleh dari hasil pengamatan dan penelitian.

c. Tinjauan Pustaka (*Research Library*)

Tinjauan pustaka dilakukan dengan cara membaca, mengutip dan membuat catatan yang bersumber pada bahan-bahan pustaka yang mendukung dan berkaitan dengan penelitian dalam hal ini mengenai data mining Naïve Bayes, K-NN.

3.2 Teknik Analisis Data

Algoritma yang digunakan dalam penelitian ini yaitu algoritma Naïve Bayes dan K-Nearest Neighbor (KNN), untuk mengklasifikasikan dalam memperkirakan waktu tunggu siswa memperoleh pekerjaan. Setelah diketahui hasil dari akurasi tiap algoritma kemudian dibandingkan untuk mencari algoritma terbaik.

a. Metode Naïve Bayes

Langkah-langkah yang dilakukan dalam mengolah data pengklasifikasian dengan metode *naive* Bayes menggunakan bantuan software *Rapid miner* dalam penelitian dapat dilihat pada Gambar 3.1 berikut ini :



Gambar 1. *Diagram Alir* Metode Naive Bayes

Berdasarkan *Diagram alir* metode *Naive Bayes* diatas terdapat :

- Input Data.
 - Menghitung nilai probabilitas awal (prior) berdasarkan Persamaan.
 - Menghitung probabilitas setiap variabel bebas pada setiap kelompok berdasarkan Persamaan.
 - Menghitung probabilitas akhir (posterior) pada masing-masing kelas berdasarkan persamaan.
 - Mencari nilai maksimum pada kedua kelompok dan menentukan hasil klasifikasi obyek.
 - Mengevaluasi hasil klasifikasi dengan menghitung nilai *precision* dan *recall*.
- b. Metode *K-Nearest Neighbors* (K-NN)
Langkah-langkah yang dilakukan dalam mengolah data pengklasifikasian dengan metode K-NN adalah sebagai berikut:



Gambar 2. *Diagram Alir* Metode K-NN

Berdasarkan *Diagram alir* metode *K-NN* diatas terdapat :

- Input Data.
- Menentukan nilai parameter K .
- Menghitung jarak Euclidean objek terhadap data training yang diberikan berdasarkan Persamaan.
- Mengurutkan data yang mempunyai jarak terkecil hingga terbesar.
- Menentukan hasil klasifikasi dengan menggunakan kategori *nearest neighbor* yang paling banyak.
- Mengevaluasi hasil klasifikasi dengan menghitung nilai *precision* dan *recall*.

4. Hasil dan Pembahasan

4.1 Data

Sebelum melakukan proses klasifikasi data, terlebih dahulu dilakukan pengumpulan data. Data yang akan diolah dalam pembahasan ini adalah data siswa di SMKS PGRI 2 Pringsewu. Data siswa yang digunakan adalah data siswa Alumni. Atribut-atribut data yang digunakan adalah Jurusan, Nilai rata-rata, Prestasi Akademik/Non-Akademik dan Ekstrakurikuler. Data diperoleh dari di SMKS PGRI 2 Pringsewu yaitu data alumni 5 tahun terakhir.

4.2 Pembagian data *training* dan data *testing*

Proses pertama sebelum melakukan analisis klasifikasi data *tracer* alumni SMKS PGRI 2 Pringsewu dengan menggunakan metode *naive* Bayes dan K-NN, data *tracer* alumni SMKS PGRI 2 Pringsewu dibagi menjadi data training dan data testing dengan proporsi pembagian 80:20.

Data training ini akan digunakan dalam proses perhitungan pada metode

naive Bayes serta perhitungan jarak pada metode K-NN, sedangkan 80 data terakhir untuk masing-masing variabel penelitian berfungsi sebagai data testing yang nantinya digunakan untuk melakukan proses klasifikasi dan pengukuran tingkat akurasi.

4.3 Analisis lama masa tunggu alumni mendapatkan pekerjaan dengan Naïve Bayes

pada metode naive Bayes dilakukan pengukuran tingkat akurasi dengan menghitung peluang kesalahan klasifikasi menggunakan precision recall dan *confusion matrix* pada persamaan adapun hasil akurasi yang didapatkan adalah sebagai berikut.

Tabel 1. Convusion Matrix Metode Naïve Bayes.

	true Cepat	true Lama	true Sedang	class precision
pred. Cepat	18	1	2	85.71%
pred. Lama	4	48	1	90.57%
pred. Sedang	2	0	4	66.67%
class recall	75.00%	97.96%	57.14%	

Hasil akurasi yang didapat kan adalah $\frac{18+48+4}{80} = 87,50 \%$.

Hasil akurasi yang didapat adalah 87,50% dimana *naïve bayes* berhasil diklasifikasi kelas cepat dengan 18 Klasifikasi benar dengan hanya 3 Klasifikasi salah, dan Klasifikasi kelas Lama 48 benar dengan hanya 5 Klasifikasi salah, sedangkan pada kelas sedang 4 kali benar dengan 2 kali salah. Dari total 80 data training pada ujicoba *naïve bayes* hanya memiliki kesalahan pada 10 Klasifikasi.

4.4 Analisis lama masa tunggu alumni mendapatkan pekerjaan dengan K-NN

Pada proses klasifikasi metode K-NN terdapat tiga alur yaitu menentukan nilai parameter K, menghitung jarak Euclid antara data training dan data testing, kemudian menentukan ranking dari hasil perhitungan jarak. Berikut adalah hasil akurasi dari setiap percobaan menggunakan nilai K yang berbeda.

Tabel 2 Hasil Percobaan perhitungan menggunakan Nilai K.

Nilai K	Akurasi
1	72.50%
3	57.50%
5	57.50%
7	78.75%
9	83.75%

Nilai K pada algoritma KNN mendefinisikan berapa banyak tetangga yang akan diperiksa untuk menentukan klasifikasi pada data tersebut. Maka perbedaan Nilai K pada setiap pengujian mempengaruhi terhadap akurasi, pada penelitian ini akan digunakan hasil nilai akurasi yang paling besar dari percobaan menggunakan 5 nilai K yang berbeda, yaitu akurasi maksimal yang di dapat adalah 83.75% dengan nilai K=9.

4.5 Perbandingan tingkat akurasi metode Naïve bayes dan K-NN

Pengukuran tingkat akurasi baik pada metode naive Bayes maupun K-NN dilakukan dengan menghitung peluang kesalahan klasifikasi. Pada proses klasifikasi diharapkan melakukan klasifikasi pada semua obyek dengan benar, sehingga semakin kecil kesalahan klasifikasi menunjukkan bahwa semakin baik hasil klasifikasi yang didapatkan. Berikut adalah hasil akurasi dari masing-masing metode :

Tabel 3 Hasil Akurasi Metode Naïve bayes dan K-NN

Metode	Akurasi
Naïve Bayes	87.50%
K-NN	83.75%

Hal ini menunjukkan bahwa metode naive Bayes bekerja lebih baik dibandingkan dengan metode K-NN untuk klasifikasi waktu tunggu alumni dalam mendapatkan pekerjaan dengan nilai akurasi yang lebih tinggi.

5. Penutup

5.1 Kesimpulan

Berdasarkan pembahasan yang telah diuraikan tentang klasifikasi waktu tunggu alumni mendapatkan pekerjaan pada SMKS PGRI 2 Pringsewu menggunakan metode *Naïve Bayes* dan metode *K-NN* pada data alumni SMKS PGRI 2 Pringsewu, maka dapat disimpulkan beberapa hal sebagai berikut:

- a. Hasil perhitungan probabilitas parameter setiap variabel bebas terhadap kelas. sangat berpengaruh terhadap hasil klasifikasi metode *Naïve bayes*, sedangkan pada metode *K-NN*, Percobaan yang dilakukan untuk mendapatkan nilai akurasi yang maksimal dapat dilakukan pada $K=1$ sampai dengan $K=9$.
- b. Jika ditinjau dari hasil akurasi yang didapatkan maka algoritma *Naïve bayes* dapat menyelesaikan studi kasus ini dengan lebih baik yaitu 87.50% tingkat akurasi, sedangkan algoritma *K-NN* Dalam percobaan $K = 9$ Menghasilkan akurasi 83.75%.

5.2 Saran

Adapun karena adanya keterbatasan dalam penelitian ini maka saran yang diajukan untuk penelitian selanjutnya adalah sebagai berikut:

1. Penelitian ini dapat dijadikan sebagai salah satu referensi bagi Sekolah Menengah Atas / Sederajat lainnya untuk digunakan sebagai acuan dalam Klasifikasi waktu tunggu alumni mendapatkan pekerjaan.
2. Dalam penelitian ini penulis hanya membahas perbandingan metode *Naïve bayes* dan *K-NN*. Oleh karena itu untuk penelitian selanjutnya penulis menyarankan agar dapat mengembangkan metode-metode analisis yang lain.

References

- [1] Ridwan, M. 2013. "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Siswa Menggunakan Algoritma Naive Bayes Classifier". *EECCIS*, 7 (1), 59-64.
- [2] Iswahyudi, C. 2015. *Sistem Pendukung Keputusan Menentukan Status Kenaikan Gaji Karyawan pada CV. Niaga Pratama Motor dengan Menggunakan Metode C4.5*. Denpasar: Program Studi Sistem Informasi STMIK STIKOM BALI.
- [3] Chala, Beyene. 2018. "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Technique". *International Journal of Pure and Applied Mathematics*
- [4] Andriani, A. 2013. "Sistem Pendukung Keputusan Berbasis Decision Tree dalam pemberian Beasiswa" dalam *Seminar Nasional Teknologi Informasi dan Komunikasi Volume 9* (pp. 163-168). Jakarta
- [5] Faridayati, L. 2014. *Rancang Bangun Sistem Pendukung Keputusan Pemberian Reward dan Punishment pada karyawan di RSNU Muslimat Jombang*. Jombang: Fakultas Teknik Sistem Informasi UNIPDU
- [6] Fowler, M. 2005. *UML Distilled 3th Ed. Panduan Singkat Bahasa Pemodelan Objek Standar*. Jogjakarta: C.V. ANDI.
- [7] Ihsan A. 2017. "Students performance prediction using KNN and Naïve Bayesian". *2017 8th International Conference on Information Technology (ICIT)*.
- [8] Puro, D. A., Andriani, K. K. W., & Fitriasih, S. S. H. 2014. "Sistem Pendukung Keputusan Penetapan Kenaikan Gaji Karyawan dengan Menggunakan Metode Algoritma C4. 5 pada PT. Harapan Makmur Surakarta" dalam *Jurnal Ilmiah SINUS*, 12 (1).
- [9] Putri, S. A. 2015. "Kombinasi Integrasi Metode Sampling dengan Naive Bayes Untuk Ketidakseimbangan Kelas Pada Klasifikasi Cacat Perangkat Lunak". dalam *Konferensi Nasional Ilmu Pengetahuan dan Teknologi (KNIT)*, 1-7.
- [10] Soebandi, R. 2012. "Pendekatan Zachman Framework untuk Sistem Informasi Administrasi Palang Merah Indonesia" dalam *Jurnal Ilmiah Ilmu Komputer*, 9 (1), 2325.
- [11] Saleh, A. 2017. "Implementasi Metode Klasifikasi Naive Bayes Dalam Diklasifikasi Besarnya Penggunaan Listrik Rumah Tangga" dalam *Citec Jurnal*, 2 (3), 207-2017.
- [12] Sardiarinto, S. 2013. "Aplikasi Sistem Pendukung Keputusan Kelayakan Peminjaman Kredit Nasabah Koperasi Berbasis Android". *Bianglala Informatika*, 1 (1).
- [13] Utami, N. K., Sukarsa, I. K., & Kencana, I. P. 2013. "Penerapan Metode Generalized Ridge Regression Dalam Mengatasi Masalah Multikolinearitas" dalam *E-Jurnal Matematika*, 2 (1), 5459.

- [14] Wasiati, H., & Wijayanti, D. 2014. "Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode Naive Bayes (Studi Kasus: PT. Karyatama Mitra Sejati Yogyakarta)" *dalam Indonesian Jurnal on Networking and Security* , 3 (2), 45-51.
- [15] BRILLIANT, Muhamad, et al. Implementation of Data Mining Using Association Rules for Transactional Data Analysis. In: *Proceeding International Conference on Information Technology and Business*. 2017. p. 177-180.
- [16] NURHASANAH, Iis Ariska, et al. Analisis Perancangan E-Business B2C (Business to Consumer) Upaya Digitalisasi Pengembangan UMKM (Studi Kasus: Kabupaten Pesawaran, Indonesia). *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, 2022, 11.3: 236-248.
- [17] HANDOKO, Dwi; LESTARI, Kanti; ASTUTI, Tiara Widi. PERANCANGAN APLIKASI PELAYANAN JASA PADA RIAS PENGANTIN (MUA) MAKHSUN PHOTOGRAPHY GISTING BERBASIS ANDROID. *SEAT: Journal Of Software Engineering and Technology*, 2021, 1.1.