

K-Nearest Neighbor Algorithm for Classification Inaccuracy in Payment of SPP Money

Algoritma K-Nearest Neighbor Untuk Klasifikasi Ketidaktepatan Waktu Pembayaran Uang SPP

Meita Dwi Anggraeni¹, Kusrini², M. Rudyanto Arief³

^{1,2,3}Teknik Informatika, Magister Teknik Informatika, Universitas Amikom Yogyakarta
E-mail: ¹meitadww@gmail.com, ²kusrini@amikom.ac.id, ³rudy@amikom.ac.id

Abstract – SMK Ma'arif Salam is one of the private educational institutions under LP Maarif. SMK Ma'arif Salam most of the school's operational financing is borne by students, especially in paying tuition fees. In general, tuition fees are paid by private schools where funding is charged to each student. Because private schools are responsible for managing education within local policies. In contrast to state schools, the costs are borne by the government/state. This research uses the K-Nearest Neighbor Algorithm. By using the K-Nearest Neighbor algorithm with a distance of K3, a pure dataset is obtained. Dividing the training data and testing data in a ratio of 80%: 20% results in 406 training data and 102 testing data. Data processing is carried out using the K-Nearest Neighbor algorithm in the validation process so that accurate results can be obtained. Calculations using the K-Nearest Neighbor algorithm obtained accuracy results of 82.35%.

Keywords — classification, k-nearest neighbor, SPP

Abstrak – SMK Ma'arif Salam merupakan salah satu lembaga pendidikan swasta dibawah LP Maarif. SMK Ma'arif Salam dalam pembiayaan operasional sekolah sebagian besar dibebankan kepada siswa, terutama dalam pembayaran SPP. Pada umumnya pembayaran SPP diterapkan sekolah swasta dimana pembiayaan dibebankan pada setiap siswa. Karena sekolah swasta bertanggung jawab mengelola pendidikan dalam kebijakan lokal. Berbeda dengan sekolah negeri, biayanya telah ditanggung oleh pemerintah/negara. Penelitian ini menggunakan Algoritma K-Nearest Neighbor. Dengan menggunakan Algoritma *K-Nearest Neighbor* dengan jarak $k=3$ sehingga didapatkan *dataset* murni. Klasifikasi dilakukan dengan membagi data training dan data testing dengan perbandingan 80% : 20% maka dihasilkan data training sebanyak 406 dan data testing sebanyak 102. Pemrosesan data dilakukan dengan Algoritma *K-Nearest Neighbor* dalam proses validasi sehingga dapat diperoleh hasil akurasi. Perhitungan dengan menggunakan algoritma *K-Nearest Neighbor* memperoleh hasil akurasi sebesar 82.35%.

Kata Kunci — k-nearest neighbor, klasifikasi, SPP

1. PENDAHULUAN

SMK Ma'arif Salam merupakan salah satu lembaga pendidikan swasta yang dikelola oleh LP Maarif. SMK Ma'arif Salam. Dalam pembiayaan operasional sekolah sebagian besar dibebankan kepada siswa, terutama dalam pembayaran Sumbangan Pembinaan Pendidikan (SPP). Pada umumnya pembayaran SPP diterapkan oleh sekolah swasta dimana pembiayaan operasional sekolah dibebankan pada setiap siswa.

Biaya pendidikan merupakan salah satu *input* yang menunjang penyelenggaraan pendidikan [1]. Biaya-biaya tersebut memegang peranan yang sangat penting dalam memperoleh pendidikan yang

berkualitas di sekolah. Dalam hal ini salah satu biaya pendidikan di SMK Ma'arif Salam adalah biaya bulanan yang dikenal dengan SPP. Hal ini karena sekolah swasta bertanggung jawab mengelola pendidikan dalam kebijakan lokal. Berbeda dengan sekolah negeri, biaya operasionalnya telah ditanggung oleh Pemerintah/Negara.

Permasalahan tersebut menjadi salah satu yang menyebabkan banyak terjadinya ketidaktepatan pembayaran SPP oleh siswa, sehingga dapat menimbulkan turunnya sumber dana yang masuk ke sekolah. Dalam hal pembayaran SPP setiap bulannya sangat diperlukan untuk kelangsungan operasional pihak sekolah, seperti pembayaran gaji guru dan karyawan, perbaikan sarana prasarana dan fasilitas proses pembelajaran. Untuk itu butuh suatu pemecahan terhadap kasus dengan cara mengklasifikasikan pembayaran SPP berdasarkan tingkat ketidaktepatan waktu pembayaran. Sehingga dapat menjadikan evaluasi bagi sekolah sesuai dengan melihat waktu yang telah ditetapkan.

Sebagai bahan acuan dan perbandingan, peneliti melakukan studi literatur dengan ulasan penelitian terdahulu yang terkait dengan metode yang pernah digunakan untuk mengklasifikasi ketidaktepatan pembayaran SPP, beberapa diantaranya pada penelitian Intan P.J Tafonao dkk. yang berjudul "Teknik Klasifikasi Prediksi Kelulusan Mahasiswa Sistem Informasi Universitas Sari Mutiara Indonesia Menggunakan *K-Nearest Neighbors*" yang bertujuan untuk meningkatkan kualitas pada perguruan tinggi Universitas Sari Mutiara Indonesia pada Fakultas Saint, Teknogi dan Informasi, Program Studi Sistem Informasi. Hasil penelitian tersebut menunjukkan hasil akurasi tertinggi menggunakan Algoritma *K-Nearest Neighbors* (KNN) adalah 95% dengan $k=3$ [2]. Selanjutnya penelitian Qurotul A'yuniyah dkk. yang berjudul "Penerapan Algoritma *K-Nearest Neighbor* Untuk Klasifikasi Jurusan Siswa Di SMA Negeri 15 Pekanbaru" yang bertujuan untuk membantu proses penjurusan siswa yang memiliki permasalahan. Hasil penelitiannya menunjukkan validasi dalam menguji prediksi dan nilai akurasi maka diperoleh hasil bahwa algoritma KNN memiliki akurasi yang tinggi dengan presentase 93.52% [3]. Selanjutnya penelitian Muhammad Rhosyid Akhmad dkk. yang berjudul "Implementasi *K-Nearest Neighbor* Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di perguruan Tinggi " yang bertujuan untuk menentukan parameter keterlambatan dan pembayaran biaya kuliah. Hasil penelitiannya menunjukkan data terbaik menunjukkan pada $k=3$ yaitu 52,82% [4]. Selanjutnya penelitian Candi Prabowo yang berjudul "Teknik Klasifikasi Pembayaran SPP Berdasarkan Tingkat Ketepatan Pembayaran" yang bertujuan Untuk mengatasi permasalahan tersebut perlu dilakukan pendeteksian faktor –faktor yang menjadi penyebab terjadinya keterlambatan pembayaran SPP dengan menggunakan *data mining*. Teknik *data mining* yang digunakan adalah klasifikasi dengan metode algoritma *Naive Bayes*. model yang dihasilkan diperoleh tingkat accuracy sebesar 63,64% [5].

Berdasarkan penelitian terdahulu maka penelitian ini bertujuan untuk mengklasifikasikan waktu pembayaran SPP berdasarkan dari ketidaktepatan waktu pembayaran dengan menggunakan algoritma *K-Nearest Neighbor* sebagai algoritma klasifikasi yang sederhana dan baik [6]. Dalam penelitian ini diharapkan dapat menyelesaikan serta mengevaluasi permasalahan mengenai ketidaktepatan waktu pembayaran SPP di SMK Ma'arif Salam.

2. METODE PENELITIAN

2.1. Klasifikasi

Menurut Dicky Nofriansyah, klasifikasi merupakan sebuah proses *training* (pembelajaran) suatu fungsi tujuan (target) yang digunakan untuk memetakan tiap himpunan atribut suatu objek ke satu dari label kelas tertentu yang didefinisikan sebelumnya [7]. Sedangkan Retno Tri Vulandari mengemukakan bahwa klasifikasi merupakan suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan [8]. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan.

2.2. *K-Nearest Neighbor*

K-Nearest Neighbour (KNN) merupakan algoritma klasifikasi klasik yang telah diterapkan di banyak bidang [9] dan menjadi salah satu teknik *lazy learning* serta termasuk



dalam kategori *instance-based learning* [10]. KNN telah banyak digunakan di beberapa bidang seperti *data mining*, pengenalan pola, dan deteksi intrusi. Prinsip kerja KNN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan *k* tetangga (*Neighbour*) terdekatnya dalam data pelatihan. Persamaan 1 merupakan urutan proses kerja KNN [11]:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots(1)$$

Keterangan:

- $d(x,y)$ = Jarak *Euclidean*
- x_i = Data *Training* ke-*i*
- y_i = Data *Testing* ke-*i*

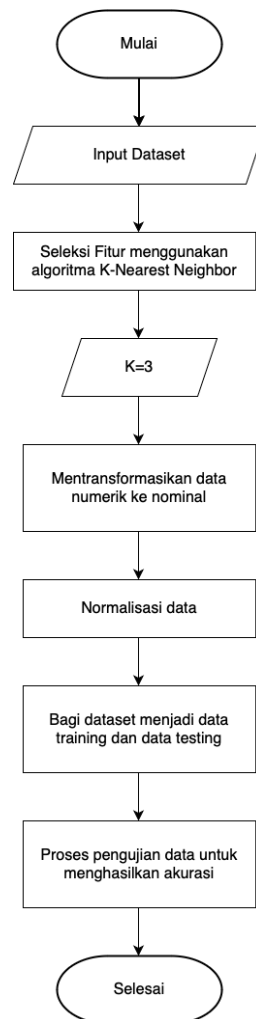
2.3. Metode Penelitian

Penelitian ini menggunakan Algoritma KNN. Pada seleksi fitur "*Input Missing Value*" dan "*Rename Unused Value*" dengan menggunakan algoritma KNN dengan jarak $k=3$ sehingga didapatkan *dataset* murni. Dalam mentransformasikan data dari numerik ke nominal dan melakukan normalisasi menentukan bentuk data yang paling tepat. Penelitian ini menggunakan *dataset* sebanyak 508. Selanjutnya *dataset* tersebut dibagi menjadi data *training* dan data *testing* dengan perbandingan 80% : 20%. Dengan perbandingan tersebut maka dihasilkan data *training* sebanyak 406 dan data *testing* sebanyak 102. Pemrosesan data dilakukan dengan menggunakan algoritma KNN untuk melakukan klasifikasi ketidaktepatan waktu pembayaran SPP di SMK Ma'arif Salam. Untuk mengetahui performa klasifikator dilakukan proses validasi sehingga dapat diperoleh hasil akurasi yang dihitung menggunakan Persamaan 2 [12], [13], [14].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \dots\dots\dots(2)$$

Dimana *TP (True Positive)* merupakan data positif yang diprediksi dengan benar, *TN (True Negative)* merupakan data negatif yang diprediksi dengan benar, *FP (False Positive)* merupakan data positif yang diprediksi dengan salah, *FN (False Negative)* merupakan data negatif yang diprediksi dengan salah.

Penelitian ini dilakukan untuk bahan pengembangan evaluasi dan pemecahan masalah terkait ketidaktepatan waktu pembayaran SPP di SMK Ma'arif Salam.



Gambar 1 Metode Alur Penelitian

3. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan data pembayaran SPP yang diperoleh dari sekolah sebagaimana disajikan oleh Tabel 1 berikut :

Tabel 1. Data Pembayaran SPP

No	Jenis Informasi	Data
1	Pekerjaan Ayah	Buruh, Karyawan Swasta, Pedagang, PNS, Tani, Wiraswasta
2	Penghasilan Ayah	<Rp. 500.000, Rp. 500.000 - Rp.1.000.000, Rp. 1.000.000 - Rp. 2.000.000, >Rp. 2.000.000
3	Pendidikan ayah	SD, SLTP, SLTA, S1, Tidak Sekolah
4	Umur Ayah	<35, 35 – 50, >50
5	Pekerjaan Ibu	Buruh, IRT, Karyawan Swasta, Pedagang, PNS, Tani, Wiraswasta
6	Penghasilan Ibu	<Rp. 500.000, Rp. 500.000 - Rp.1.000.000, Rp. 1.000.000 - Rp. 2.000.000, >Rp. 2.000.000
7	Pendidikan Ibu	SD, SLTP, SLTA, S1, Tidak Sekolah
8	Umur Ibu	<35, 35 – 50, >50
9	Tanggungjawab Keluarga	>2, <2

Hasil klasifikasi yang diperoleh diambil dari data pada Tabel 1 dan dilakukan klasifikasi menggunakan Algoritma KNN disajikan oleh Tabel 2 sebagai berikut :

Tabel 2. Perhitungan dengan *K-Nearest Neighbor*

Accuracy : 82.35%

	<i>True TEPAT</i>	<i>True TIDAK TEPAT</i>	<i>class precision</i>
pred. TEPAT	22	6	78.57%
pred. TIDAK TEPAT	12	62	83.78%
<i>class recall</i>	64.71%	91.18%	

Klasifikasi ketidaktepatan waktu pembayaran SPP di SMK Ma'arif Salam dengan menggunakan algoritma *K-Nearest Neighbor* memperoleh hasil akurasi sebesar 82.35%. dengan akurasi tersebut dapat disimpulkan bahwa bahwa KNN dapat digunakan dalam mengetahui ketidaktepatan pembayaran uang SPP di SMK Ma'arif Salam. Hasil akurasi tersebut diperoleh dengan menggunakan Persamaan 2.

4. KESIMPULAN

Berdasarkan dari hasil penelitian penggunaan Algoritma KNN untuk klasifikasi ketidaktepatan waktu pembayaran uang SPP di SMK Ma'arif Salam yang telah dilakukan dapat disimpulkan bahwa penelitian ini dapat memperoleh hasil persentase akurasi sebesar 82.35% dengan perbandingan data *training* dan data *testing* sebesar 80% : 20%. Dengan akurasi yang didapatkan sebesar 82,35 menunjukkan bahwa KNN dapat digunakan dalam mengetahui ketidaktepatan pembayaran uang SPP di SMK Ma'arif Salam.

DAFTAR PUSTAKA

- [1] S. Sudarmono, L. Hasibuan, dan K. A. Us, "Pembiayaan Pendidikan," *Jurnal Manajemen Pendidikan Dan Ilmu Sosial*, vol. 2, no. 1, hlm. 266–280, 2020.
- [2] I. P. J. Tafonao dan A. F. K. Sibero*, "Teknik Klasifikasi Prediksi Kelulusan Mahasiswa Sistem Informasi Universitas Sari Mutiara Indonesia Menggunakan K- Nearest Neighbors," *JURNAL MAHAJANA INFORMASI*, vol. 7, no. 1, 2022, doi: 10.51544/jurnalmi.v7i1.2956.
- [3] Q. A. A'yuniyah dan M. Reza, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru," *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, vol. 3, no. 1, 2023, doi: 10.57152/ijirse.v3i1.484.
- [4] M. R. Akhmad dan T. A. Y. Siswa, "Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi," *Progresif: Jurnal Ilmiah Komputer*, vol. 18, no. 2, 2022, doi: 10.35889/progresif.v18i2.921.
- [5] C. Prabowo, L. Nul, H. Cep, L. Rohmat, dan A. Rizki, "Teknik Klasifikasi Pembayaran SPP Berdasarkan Tingkat Ketepatan Pembayaran," *Jurnal Data Science & Informatika*, vol. 1, no. 1, 2021.
- [6] N. Sari dan R. Wulanningrum, "Implementation of the K-Nearest Neighbor Algorithm for Identification of Orchid Flower Image," *JTECS : Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem dan Komputer*, vol. 1, no. 2, 2021, doi: 10.32503/jtecs.v1i2.1750.

- [7] D. Nofriansyah dan G. W. Nurcahyo, "Algoritma Data Mining Dan Pengujian," *Algoritma Data Mining dan Pengujian*. 2019.
- [8] R. T. Vuldari, "Data Mining Teori dan Aplikasi Rapidminer," *Yogyakarta: Gava Media*, 2017.
- [9] Y. Wang, Z. Pan, dan J. Dong, "A new two-layer nearest neighbor selection method for kNN classifier," *Knowl Based Syst*, vol. 235, 2022, doi: 10.1016/j.knosys.2021.107604.
- [10] A. Prabowo, D. Erwanto, dan P. N. Rahayu, "Klasifikasi Kesegaran Daging Sapi Menggunakan Metode Ekstraksi Tekstur GLCM dan KNN Freshness Classification of Beef Using GLCM Texture Extraction Method and KNN," *Jec*, vol. 7, no. 1, 2021.
- [11] F. Gorunescu, "Data mining: Concepts, models and techniques," *Intelligent Systems Reference Library*, vol. 12, 2011, doi: 10.1007/978-3-642-19721-5.
- [12] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, 2020.
- [13] T. Sutojo, P. S. Tirajani, C. A. Sari, E. H. Rachmawanto, dan others, "CBIR for classification of cow types using GLCM and color features extraction," dalam *2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2017, hlm. 182–187.
- [14] D. Chicco dan G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, hlm. 1–13, 2020.