



PEMILIHAN FITUR DENGAN FORWARD SELECTION PADA METODE K-NEAREST NEIGHBOR UNTUK PREDIKSI BIMBINGAN KONSELING SISWA

Wildani Eko Nugroho

Politeknik Harapan Bersama Tegal
wild4n1@gmail.com

ABSTRACT

One of the most difficult issues to predict in schools is that of guidance and counseling. For individuals who have the authority to make decisions, accurate predictions are needed to support decision-making. KNN is one of the methods that can make predictions, especially for counseling guidance predictions. The KNN algorithm approach used to conduct tests and experiments on 388 student data obtained an accuracy of 96.40%. After optimization using feature selection with forward selection, the accuracy increased by 0.51%. The final result of the accuracy obtained after experiencing an increase in the level of accuracy value of 96.91% with the number of students who do not have to do counseling guidance is as many as 362 students, and 62 students must do counseling guidance.

Keywords : . KNN, Feature Selection, Forward Selection, Prediction, Guidance, Counseling

ABSTRAK

Salah satu masalah yang sulit untuk diramalkan yang ada di sekolah adalah masalah bimbingan dan konseling. Bagi individu yang memiliki wewenang untuk mengambil keputusan, prediksi yang akurat diperlukan untuk mendukung pengambilan keputusan. KNN merupakan salah satu metode yang dapat melakukan prediksi terutama untuk prediksi bimbingan konseling. Pendekatan algoritma KNN yang digunakan untuk melakukan pengujian dan eksperimen terhadap 388 data siswa, memperoleh hasil akurasi sebanyak 96.40%. Setelah dilakukan optimasi menggunakan pemilihan fitur dengan forward selection mengalami kenaikan akurasi sebesar 0.51%. Hasil akhir dari akurasi yang diperoleh setelah mengalami kenaikan tingkat nilai akurasi sebesar 96.91% dengan jumlah siswa yang tidak harus melakukan bimbingan konseling sebanyak 362 siswa dan 62 siswa harus melakukan bimbingan konseling.

Kata Kunci: KNN, Feature Selection, Forward Selection, Prediksi, Bimbingan, Konseling.

1. PENDAHULUAN

Bimbingan dan konseling yang sesuai dengan paradigma perkembangan (kuratif) tidak mengabaikan layanan yang berfokus pada pencegahan masalah (relief) (Susanto 2018). Potensi (kecerdasan, kemampuan, hobi, kepribadian, kondisi fisik, riwayat keluarga, dan latar belakang pendidikan) setiap siswa dan konseling berbeda-beda. Oleh karena itu, konselor dan siswa membutuhkan berbagai jenis layanan perkembangan siswa yang berbeda-beda (Han and Goleman, Daniel; Boyatzis, Richard; McKee 2019).

Bimbingan untuk siswa dipengaruhi oleh aspek lingkungan, fisik, psikologis, dan sosial. Sesuai dengan sifatnya, lingkungan berubah

secara terus – menerus. Gaya hidup masyarakat, terutama siswa dan konselor, mungkin terpengaruh oleh perubahan lingkungan (Hifsy, Hariko, and Karneli 2022). Siswa yang menerima bimbingan pada dasarnya mampu menyesuaikan diri dengan lingkungan dan dirinya sendiri. Guru pembimbing atau guru yang bertugas memberikan layanan bimbingan dan konseling di sekolah bertugas membantu siswa baik secara individu maupun kelompok agar dapat mandiri dan berkembang secara optimal dalam bidang bimbingan pribadi, sosial. Hal seperti ini dapat dilakukan melalui berbagai kegiatan yang bersifat positif berdasarkan norma yang berlaku dalam lingkungan (Ramlah 2018).

Salah satu metode yang dapat digunakan untuk melakukan analisis sebuah prediksi yaitu data mining. Metode ini diharapkan dapat menyelesaikan masalah prediksi tentang bimbingan konseling siswa. Machine Learning merupakan salah satu teknologi yang saat ini sangat tepat untuk diterapkan pada permasalahan prediksi terutama prediksi tentang bimbingan konseling. Seperti halnya neural network, support vector machine, deep learning, KNN dan metode lainnya dalam berbagai bidang penerapannya seperti pendidikan dan kesehatan (Somantri, Nugroho, and Supriyono 2022).

Pada penelitian terdahulu terkait dengan analisis data bimbingan konseling seperti halnya penelitian tentang metode algoritma naive bayes berbasis forward selection untuk prediksi bimbingan konseling siswa (Fanani 2020a). Penelitian lainnya adalah dilakukan penggabungan metode forward selection untuk pemilihan fitur pada metode naive bayes (Fanani 2020b).

Penelitian selanjutnya terkait dengan bimbingan konseling yaitu menentukan prediksi rekomendasi bimbingan konseling siswa menengah kejuruan menggunakan metode KNN (W. E. Nugroho and Saputro 2022). KNN merupakan salah satu metode yang berkembang dengan kelebihan dapat diterapkan dan melakukan prediksi bimbingan konseling untuk siswa.

Dalam artikel ini, penelitian yang dilakukan bagaimana metode KNN memerlukan optimalisasi, karena performance akurasi yang dihasilkan masih perlu ditingkatkan dan masih mempunyai hasil tingkat akurasi yang belum memuaskan. Selain itu, pada langkah preprocessing data masih terdapat masalah pada saat pemilihan fitur yang masih belum optimal. Permasalahan tersebut dapat mempengaruhi hasil akurasi pada metode yang digunakan. Untuk mengatasi masalah ini, maka pada penelitian ini dipilihlah metode optimasi dengan feature selection yang dalam hal ini menggunakan metode forward selection. Tujuan dari penelitian ini adalah mencari model terbaik untuk prediksi bimbingan konseling siswa dengan menggunakan metode KNN dengan optimalisasi menggunakan metode forward selection.

2. KERANGKA TEORI

2.1. K-Nearest Neighbor (KNN)

Algoritma KNN adalah metode yang menggunakan algoritma supervised. Dengan mencampurkan pola yang telah ditemukan sebelumnya dengan data baru, pembelajaran supervised, yang berlawanan dengan pembelajaran unsupervised, dan kemudian

mencari pola-pola baru di dalam data (Gunawan, Zarlis, and Roslina 2021).

Sedangkan tujuan dari pembelajaran unsupervised adalah menentukan pola dalam data, data dalam pembelajaran supervised sudah mengandung pola. Tujuan KNN yaitu mengategorikan objek baru berdasarkan properti dan data training (Kustiyahningsih, Anamisa, and Syafa'ah 2013).

Dari penjelasan diatas maka dapat dijabarkan dengan rumus perhitungan sebagai berikut:

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (1)$$

Dimana : x_1 = data sampel, x_2 = testing data,
 i = data variabel, d = jarak, p = dimensi data

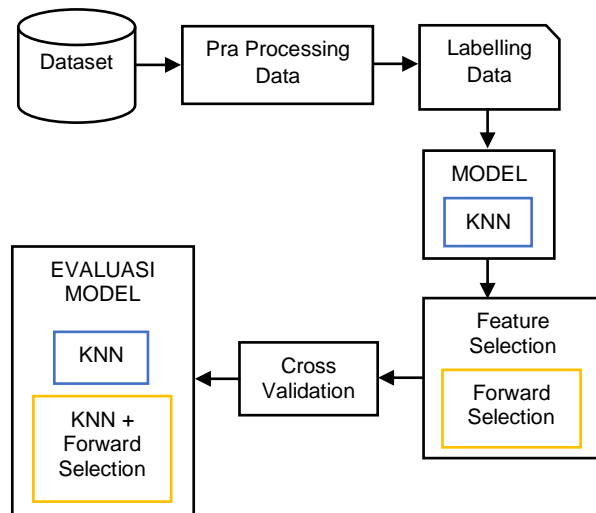
2.2. Feature Selection dan Forward Selection

Salah satu metode yang digunakan untuk mengidentifikasi properti paling penting dari sebuah dataset adalah melalui seleksi fitur. Subset terbaik dari kumpulan fitur asli harus dipilih karena tidak semua fitur atau atribut yang relevan terdapat masalah (M. F. Nugroho and Wibowo 2017). Pada kenyataannya, beberapa karakteristik atau sifat ini dapat mengganggu akurasi. Untuk meningkatkan akurasi, maka fitur yang tidak relevan dan tidak dibutuhkan akan dihapus. Selain itu, proses komputasi akan lambat oleh banyaknya fitur itu sendiri (Wang 2012).

Forward selection atau metode seleksi maju adalah algoritma pencarian yang sangat sederhana. Forward selection merupakan salah metode yang didasarkan pada metode regresi linear. Forward selection merupakan salah satu metode untuk mengurangi kompleksitas dataset dengan menghapus atribut yang tidak berguna atau berlebihan (M. F. Nugroho and Wibowo 2017). Dalam pendekatan forward selection ini, variabel ditambahkan satu per satu hingga kondisi yang telah ditentukan terpenuhi setelah dimulai dengan model yang kosong (Mahfudh and Mustofa 2019).

3. METODE

Metode yang digunakan dalam penelitian ini bisa dilihat pada bagan kerangka penelitian dibawah ini:



Gambar 1. Kerangka Penelitian

Gambar 1 diatas merupakan kerangka penelitian yang diusulkan. Data yang digunakan adalah data tahun 2021 yang berasal dari penilaian terhadap siswa di sekolah yang dilakukan oleh guru bimbingan konseling (BK).

Pada penelitian ini jumlah data yang digunakan sebanyak 388 record data dengan 2 jenis label yaitu yang konseling dan tidak di konseling. Untuk memperoleh data yang diinginkan, maka langkah awal dilakukan preprocessing data yaitu dengan cara cleaning data, menghapus atribut yang tidak relevan dan pemberian label data.

Setelah melalui proses preprocessing data langkah selanjutnya menerapkan metode algoritma KNN dan penerapan feature selection yang dalam hal ini menggunakan metode forward selection. Selanjutnya melakukan validasi data menggunakan cross validation dan evaluasi model yang digunakan. Dari hal tersebut kemudian pada tahapan evaluasi model dilakukan perbandingan model antara KNN dan KNN yang dioptimalisasi dengan feature selection yang dalam hal ini menggunakan forward selection.

4. HASIL DAN PEMBAHASAN

Tahapan selanjutnya dalam penelitian ini adalah menyampaikan hasil eksperimen, sebagai berikut:

4.1. Penerapan K-Nearest Neighbor (KNN)

Penerapan metode yang diusulkan adalah dengan metode KNN. Pada metode KNN sebelum pengujian terlebih dahulu dilakukan pengaturan parameter, yaitu untuk sampling type menggunakan Stratified Sampling, Linier Sampling dan Shuffled Sampling, sedangkan number of fold 6-10, adapun hasil pengujian dapat dilihat pada tabel 1, tabel 2, tabel 3.

Tabel 1. KNN Dengan Stratified Sampling

Sampling Type	Number Of Fold	Accuracy	Precision	Recall
Stratified Sampling	10	95.89%	70.00%	48.33%
	9	95.36%	63.64%	33.33%
	8	96.40%	75.00%	60.42%
	7	95.10%	57.14%	38.10%
	6	95.62%	69.44%	44.44%

Tabel 2. KNN Dengan Linier Sampling

Sampling Type	Number Of Fold	Accuracy	Precision	Recall
Linier Sampling	10	94.10%	40.00%	19.05%
	9	94.34%	44.44%	19.05%
	8	94.60%	50.00%	28.57%
	7	94.84%	71.43%	28.57%
	6	94.10%	37.50%	14.29%
	5	93.83%	33.33%	14.29%
	4	94.07%	56.25%	21.02%
	3	94.08%	72.22%	16.92%
	2	94.85%	58.33%	31.67%

Tabel 3. KNN Dengan Shuffled Sampling

Sampling Type	Number Of Fold	Accuracy	Precision	Recall
Shuffled Sampling	10	95.62%	62.50%	47.62%
	9	95.87%	64.71%	52.38%
	8	96.13%	66.67%	57.14%
	7	95.86%	64.71%	52.38%
	6	95.63%	62.50%	49.21%

Pada tabel 1 metode KNN dengan tingkat akurasi paling tinggi sebesar 96.40%. Hal ini merupakan model yang memiliki performace terbaik dibandingkan dengan model lain yang memiliki nilai precision 75.00% dan recall 60.42%. Dibandingkan dengan tabel 2 dan tabel 3 yang memiliki nilai akurasi, precision dan recall dibawah tabel 1. Adapun nilai akurasi paling tinggi untuk tabel 2 sebesar 94.85%, precision 58.33 dan recall 31.67%. Sedangkan untuk tabel 3 memiliki nilai akurasi sebesar 96.13%, precision 62.50%, dan recall 49.21%.

4.2. Penerapan KNN Dan Forward Selection

Setelah pengujian menggunakan metode KNN, selanjutnya menerapkan metode optimalisasi forward selection pada KNN dengan menetapkan parameter yang terbaik pada KNN. Percobaan yang dilakukan menggunakan number of fold yang dijadikan parameter terbaik. Untuk tabel 4 dan tabel 5 menggunakan number of fold 5-10, pada tabel 6 number of fold yang digunakan 6-10. Hasil pengujian forward selection pada KNN secara lengkap dapat dilihat pada tabel di bawah ini.

Tabel 4. Penerapan Forward Selection pada KNN Dengan Stratified Sampling

Sampling Type	Number Of Fold	Accuracy	Precision	Recall
Stratified Sampling	10	94.60%	50.00%	10.00%
	9	95.36%	58.82%	44.44%
	8	96.66%	83.75%	62.50%
	7	96.91%	82.38%	61.90%

6	96.13%	71.43%	48.61%
5	95.87%	100.00%	23.00%

Tabel 5. Penerapan Forward Selection pada KNN Dengan Linier Sampling

Sampling Type	Number Of Fold	Accuracy	Precision	Recall
Linier Sampling	10	95.90%	72.73%	38.10%
	9	96.14%	80.00%	38.10%
	8	95.89%	66.67%	47.62%
	7	96.39%	85.71%	52.38%
	6	94.87%	57.14%	19.05%
5	95.63%	80.00%	33.33%	

Tabel 6. Penerapan Forward Selection pada KNN Dengan Shuffled Sampling

Sampling Type	Number Of Fold	Accuracy	Precision	Recall
Shuffled Sampling	10	95.10%	66.67%	18.33%
	9	96.14%	80.00%	42.04%
	8	96.14%	65.00%	61.90%
	7	95.36%	71.43%	23.81%
	6	94.86%	58.33%	45.56%

Tabel diatas menunjukkan bahwa hasil pengujian memperoleh nilai performace terbaik pada tabel 4 dengan nilai akurasi tertinggi sebesar 96.91% dengan menggunakan parameter Stratified Sampling, number of fold 7, dan hasil precision 82.38% dan recall 61.90%.

Pada tabel 5 untuk nilai akurasinya sebesar 96.39% precision 85.71% dan recall 52.38%. sedangkan pada tabel 6 nilai akurasi yang dihasilkan 96.14% precision 85.71% dan recall 61.90%.

4.3. Pembahasan

Jika dibandingkan antara penelitian terdahulu dengan penelitian yang sedang dikerjakan maka hasilnya lebih baik dari hasil penelitian terdahulu dibandingkan dengan sekarang, perbandingan hasil penelitian terdahulu dengan sekarang dapat dilihat pada tabel 7 di bawah ini.

Tabel 7. Perbandingan Hasil Pengujian

Metode	Hasil Akurasi
KNN	96.40%
KNN dan Forward Selection	96.91%

Dari tabel diatas menunjukkan bahwa hasil penelitian terdahulu lebih baik dari penelitian yang berjalan sekarang. Hal ini karena pada penelitian terdahulu memperoleh hasil akurasi sebesar 96.40% dengan metode KNN, sedangkan setelah metode KNN dioptimasi dengan metode forward selection mengalami kenaikan tingkat akurasi sebanyak 0.51% dengan menghasilkan nilai akurasi sebesar 96.91%.

Dari pengujian diatas menunjukkan bahwa melakukan prediksi bimbingan konseling

menggunakan metode KNN dan metode KNN dengan optimasi forward selection yang lebih baik adalah metode KNN dengan optimasi forward selection dengan memperoleh nilai akurasi sebesar 96.91%.

Untuk itu, pengujian pada penelitian tentang pemilihan fitur dengan forward selection pada metode KNN untuk prediksi bimbingan konseling siswa memperoleh nilai akurasi tertinggi sebesar 96.91% dengan jumlah siswa yang tidak harus bimbingan konseling sebanyak 362 siswa dan yang harus melakukan bimbingan konseling sebanyak 26 siswa.

5. SIMPULAN

Dari eksperimen dan pengujian yang sudah dilakukan pada pemilihan fitur dengan forward selection pada metode KNN untuk prediksi bimbingan konseling siswa memperoleh hasil tingkat akurasi 96.91% dengan jumlah siswa yang tidak harus bimbingan konseling sebanyak 362 siswa dan yang harus melakukan bimbingan konseling sebanyak 26 siswa.

Hasil akurasi yang diperoleh pada penelitian yang sedang berjalan mengalami kenaikan nilai akurasinya sebanyak 0.51%. karena hasil akurasi pada penelitian terdahulu hanya memperoleh hasil akurasi sebesar 96.40%.

UCAPAN TERIMA KASIH

Ucapan terimakasih yang tidak terhingga kepada semua pihak yang telah membantu dan mendukung proses penelitian ini. Ucapan terimakasih khususnya pada pihak institusi yang telah mendukung aktifitas penelitian ini

DAFTAR PUSTAKA

- Fanani, M. Rudi. 2020a. "Algoritma Naïve Bayes Berbasis Forward Selection Untuk Prediksi Bimbingan Konseling Siswa." *Jurnal DISPROTEK* 11(1): 13–22.
- Fanani, M Rudi. 2020b. "PENGABUNGAN FORWARD SELECTION UNTUK PEMILIHAN FITUR PADA P-ISSN : 2089-676X." 9(2): 85–88.
- Gunawan, Muhammad, Muhammad Zarlis, and Roslina Roslina. 2021. "Analisis Komparasi Algoritma Naïve Bayes Dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu." *Jurnal Media Informatika Budidarma* 5(2): 513.
- Han, Eunice S., and Annie goleman, daniel; boyatzis, Richard; Mckee. 2019. "BIMBINGAN DAN KONSELING Prespektif Sekolah." *Journal of Chemical Information and Modeling* 53(9): 1689–99.
- Hifsy, Ifrah, Rezki Hariko, and Yeni Karneli. 2022. "Menciptakan Konseling Yang Kondusif Melalui Teknik-Teknik Dasar

- Konseling (Attending, Listening Dan Structuring).” *Ristekdik : Jurnal Bimbingan dan Konseling* 7(2): 143.
- Kustiyahningsih, Yeni, Devie Rosa Anamisa, and Nikmatus Syafa’ah. 2013. “Siswa Sma Menggunakan Metode Knn Dan Smart.”
- Mahfudh, Adzhal Arwani, and Hery Mustofa. 2019. “Klasifikasi Pemahaman Santri Dalam Pembelajaran Kitab Kuning Menggunakan Algoritma Naive Bayes Berbasis Forward Selection.” *Walisongo Journal of Information Technology* 1(2): 101.
- Nugroho, Mohamad Fajarianditya, and Setyoningsih Wibowo. 2017. “Fitur Seleksi Forward Selection Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes.” *Jurnal Informatika Upgris* 3(1): 63–70.
- Nugroho, W E, and H Saputro. 2022. “Menentukan Prediksi Rekomendasi Bimbingan Konseling Siswa Sekolah Menengah Kejuruan Menggunakan K-Nearest Neighbor.” *Journal of Information System and ...* 2(2): 20–24. <https://journal.unisnu.ac.id/JISTER/article/view/409>.
- Ramlah. 2018. “Pentingnya Layanan Bimbingan Konseling Bagi Peserta Didik.” *Al-Mau’lzhah* 1(September): 70–76.
- Somantri, Oman, Wildani, Eko Nugroho, and Abdul, Rohman Supriyono. 2022. “Penerapan Feature Selection Pada Algoritma Decision Tree Untuk Menentukan Pola Rekomendasi Dini Konseling.” *Jurnal Sistem Komputer dan Informatika (JSON)* 4(2): 272–79.
- Susanto, Ahmad. 2018. *Bimbingan Dan Konseling. Konsep, Teori, Dan Aplikasinya*.
- Wang, Ruihu. 2012. “AdaBoost for Feature Selection, Classification and Its Relation with SVM, A Review.” *Physics Procedia* 25: 800–807. <http://linkinghub.elsevier.com/retrieve/pii/S1875389212005767> (May 27, 2013).