

Implementasi Algoritma KNN Untuk Klasifikasi Jurusan Siswa Kelas XI di SMA Al Irsyad Tegal

Mohamad Arjun Rizqian¹ Bambang Irawan² Puji Wahyuningsih³

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhadi Setiabudi, Kabupaten Brebes, Provinsi Jawa Tengah, Indonesia^{1,2,3}

Email: mohamadarjunrizqian@gmail.com¹ bambangumus@gmail.com²
pujiwahyuningsih051087@gmail.com³

Abstrak

Penjurusan merupakan suatu proses yang dilakukan sekolah untuk memetakan minat dan bakat siswa agar sesuai dengan kemampuan yang dimiliki. Berdasarkan hasil pengamatan yang dilakukan di SMA Al Irsyad Tegal, proses penjurusan siswa memiliki kendala seperti proses yang dilakukan masih menggunakan manual yaitu dengan memetakan siswa satu persatu berdasarkan kriteria yang dibuat oleh sekolah dan masih terdapat hasil pemetaan yang tidak sesuai dengan minat dan bakat siswa. *Data Mining* memiliki suatu teknik untuk klasifikasi suatu data untuk mempermudah dalam hal penjurusan siswa. Oleh karena itu, dilakukanlah proses penjurusan siswa dengan teknik klasifikasi menggunakan *K-Nearest Neighbor* dengan menggunakan aplikasi *RapidMiner*. Hasil penelitian yang didapatkan adalah pengujian *split ratio* tertinggi dengan pembagian 80% data *training* dan 20% data *testing* diperoleh pada K=9 dengan akurasi sebanyak 77,78%, *class recall* IPA 91,67%, *class recall* IPS 50%, *precision* IPA 78,57% dan *precision* IPS 75%. Dengan hasil tersebut artinya klasifikasi jurusan dapat menggunakan metode *K-Nearest Neighbor* karena menghasilkan nilai akurasi yang baik.

Kata Kunci: Klasifikasi, Penjurusan Siswa, *K-Nearest Neighbor*

Abstract

Majoring is a process carried out by schools to map students' interests and talents so that they match their abilities. Based on the results of observations made at Al Irsyad Tegal High School, the student majoring process has problems such as the process is still manual, namely by mapping students one by one based on criteria created by the school and there are still mapping results that do not match the students' interests and talents. Data Mining has a technique for classifying data to make it easier for students to major. Therefore, the student majoring process was carried out using the classification technique using K-Nearest Neighbor using the RapidMiner application. The research results obtained were the highest split ratio testing with a division of 80% training data and 20% testing data obtained at K=9 with an accuracy of 77.78%, IPA class recall 91.67%, IPS class recall 50%, IPA precision 78.57% and IPS precision 75%. These results mean that major classification can use the K-Nearest Neighbor method because it produces good accuracy values.

Keywords: Classification, Student Majoring, *K-Nearest Neighbor*



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

PENDAHULUAN

Sekolah merupakan tempat untuk dijadikan wadah bagi siswa yang akan menimba suatu ilmu. Setiap sekolah pastinya mempunyai proses pembelajaran yang berbeda beda. Dengan adanya sekolah, setiap siswa dapat memperoleh suatu pelajaran sesuai dengan minat belajarnya. Hal itu dapat menciptakan suatu prestasi karena proses belajar yang dilakukan sesuai dengan kemampuan yang dimiliki sehingga siswa dengan mudah menyerap ilmu yang diberikan oleh guru. Dalam mewujudkan suatu prestasi bagi siswa, pastinya sekolah akan memfokuskan atau membagi jurusan sesuai dengan minat belajarnya masing masing. Seperti halnya sekolah SMA Al Irsyad Tegal yang terletak di jalan Gajah Mada Kota Tegal ini melakukan

penjurusan siswa kelas X naik ke kelas XI. Proses penjurusan siswa di SMA Al Irsyad masih menggunakan manual yaitu dengan berlandaskan evaluasi beberapa nilai akademik yang berkaitan dengan nilai IPA dan nilai IPS kelas X dan hasil tes psikologi yang di dapat dari hasil kerja sama antara SMA Al Irsyad Tegal dengan lembaga tes psikologi. Proses tersebut tentunya membutuhkan waktu yang lama karena tim BK yang berkerjasama dengan tim kurikulum harus melihat nilai akademik yang berkaitan dengan nilai IPA dan nilai IPS satu persatu dan menunggu hasil tes psikologi yang dilakukan oleh lembaga tersebut.

Selain itu, permasalahan yang biasanya terjadi pada banyak siswa saat penjurusan yaitu secara nilai akademik dan hasil tes psikologi berbeda kategori. Contohnya nilai akademik yang berkaitan dengan nilai IPA memiliki nilai yang tinggi dan nilai IPS memiliki nilai yang kurang, namun secara hasil tes psikologi siswa tersebut dikategorikan masuk kedalam nilai IPS karena memiliki nilai verbal yang tinggi sehingga tim BK harus melihat kembali satu persatu dan menanyakan ulang kepada siswa tersebut tentang minat yang akan diambil dan akan dirapatkan kembali oleh tim BK itu sendiri. Jika terjadi ketidaksesuaian penjurusan siswa akan berdampak kepada siswa yang mengakibatkan proses belajar mengajar tidak kondusif karena siswa tersebut tidak bisa mengikuti pelajaran yang ada dan akan merasa tertinggal dengan teman temannya dan juga akan menurunkan capaian prestasi akademik. Oleh karena itu dibutuhkan suatu teknik klasifikasi untuk pemilihan jurusan siswa. Peranan *Data Mining* sangat diperlukan untuk proses pemilihan jurusan siswa yang ada di SMA Al Irsyad Tegal.

Data Mining sendiri adalah serangkaian proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data dengan ukuran yang cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data tersebut. Teknik *Data Mining* salah satunya adalah *K-Nearest Neighbor*. *K-Nearest Neighbor* adalah metode klasifikasi dengan cara mencari nilai K pada data *training* ke dalam data baru atau data *testing* yang mirip atau jarak terdekatnya. Berdasarkan hasil dari penelitian yang terkait penulis akan menggunakan teknik klasifikasi dari metode algoritma *K-Nearest Neighbor* untuk klasifikasi penentuan jurusan siswa karena mengacu penelitian terkait bahwasannya metode algoritma *K-Nearest Neighbor* memiliki tingkat akurasi yang cukup tinggi. Oleh karena itu penulis memutuskan mengangkat judul "Penerapan Algoritma *K-Nearest Neighbor* Untuk Klasifikasi Jurusan Siswa Kelas XI di SMA Al Irsyad Tegal".

Tinjauan Pustaka

Data Mining

Data Mining adalah suatu proses untuk mencari nilai pengetahuan dari suatu kumpulan data yang tidak diketahui secara manual. *Data Mining* atau yang dikenal dengan istilah *Knowledge Discovery in Database (KDD)* adalah mencari suatu data untuk membangun sebuah model dan model tersebut agar dapat mengenali pola data yang lain yang tidak ada dalam *database*. *Data Mining* memiliki banyak fungsi, salah satunya fungsi deskripsi dan prediksi. fungsi deskripsi dalam *Data Mining* adalah suatu fungsi untuk mengamati data secara mendalam pada data yang diamati. Dengan adanya suatu proses tersebut, diharapkan dapat mengetahui suatu perilaku atau pola pada sebuah data. Dengan kata lain, jika pola yang ditemukan berulang dan bernilai menandakan terdapat karakteristik pada suatu data tersebut sedangkan untuk fungsi prediksi sendiri merupakan sebuah fungsi bagaimana sebuah proses nantinya dapat menemukan suatu pola tertentu dari suatu data. Suatu pola dapat dilihat dari berbagai macam variabel yang ada pada data. Saat sudah mendapatkan pola, maka pola tersebut bisa dimanfaatkan untuk memprediksi variabel yang lain yang belum diketahui nilai ataupun jenisnya. Pada saat proses pengumpulan informasi tentunya menggunakan suatu metode dan metode itu yang nantinya berguna untuk menemukan suatu data.

K-Nearest Neighbor

K-Nearest Neighbor atau *KNN* adalah metode klasifikasi dengan cara mencari nilai *K* pada data *training* ke dalam data baru atau data *testing* yang mirip atau jarak terdekatnya. Pada saat menentukan jarak terdekat, terlebih dahulu data dibagi menjadi data *training* dan data *testing*, selanjutnya di hitung jarak masing-masing data *testing* (*Euclidian Distance*) terhadap data *training*. *K-Nearest Neighbor* termasuk dalam algoritma *supervised learning*. Kategori *K-Nearest Neighbor* diklasifikasikan berdasarkan mayoritas terhadap data yang baru. *KNN* merupakan suatu metode *lazy learning* dimana model yang dipelajari dari data *testing* tidak ada, hanya berdasarkan dari contoh uji klasifikasi.

Klasifikasi

Klasifikasi merupakan salah satu tugas yang penting dalam Data Mining karena berfungsi untuk mengelompokkan suatu data ke dalam kelas-kelas yang berbeda. Definisi pengklasifikasian adalah suatu fungsi yang bersifat prediksi dan penggolongan data tertentu ke dalam sebuah kelas. Proses pengklasifikasian dibuat dari sekumpulan data latih dengan kelas yang telah ditentukan sebelumnya. Kinerja pengklasifikasian data diukur dengan ketepatan.

RapidMiner

RapidMiner merupakan suatu perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Technology Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan menggunakan tampilan *GUI* (*Graphical User Interface*) sehingga memudahkan para pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak termasuk dalam kategori *open source* dan dibuat dengan menggunakan bahasa pemrograman Java di bawah lisensi *GNU Public Licence*. *RapidMiner* dapat dijalankan pada sistem operasi manapun. Dengan menggunakan *RapidMiner*, tidak membutuhkan kemampuan koding khusus, karena semua fasilitas sudah tersedia. Aplikasi *RapidMiner* dikhususkan untuk penggunaan *Data Mining*.

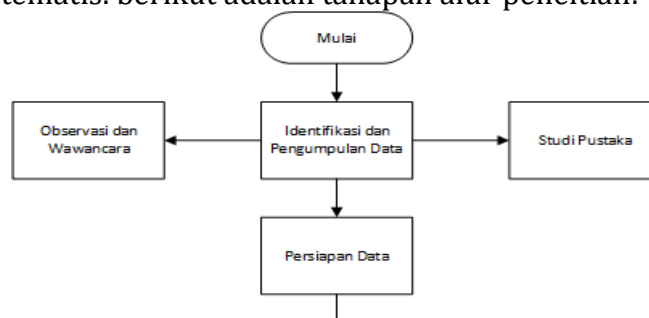
Cross Validation

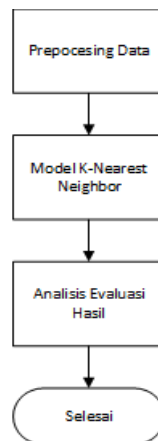
K-Fold Cross Validation merupakan suatu cara untuk mengevaluasi kinerja model pelatihan yang dibuat. *Cross Validation* merupakan metode statistik silang yang dapat digunakan untuk menguji validasi kinerja suatu model agar tidak terjadi tumpang tindih pada data uji. Proses awal pada metode ini yaitu dataset yang digunakan dibagi terlebih dahulu menjadi *K* partisi secara acak. Setiap partisi akan digunakan sebagai data *testing* satu kali, sementara partisi yang lain digunakan sebagai data *training*. Proses berulang sebanyak *K* kali, di mana setiap kali percobaan menggunakan partisi yang berbeda sebagai data *testing* [9].

METODE PENELITIAN

Diagram Alur Penelitian

Adapun beberapa tahapan yang digunakan untuk penelitian ini supaya terciptanya penelitian baik dan sistematis. berikut adalah tahapan alur penelitian:

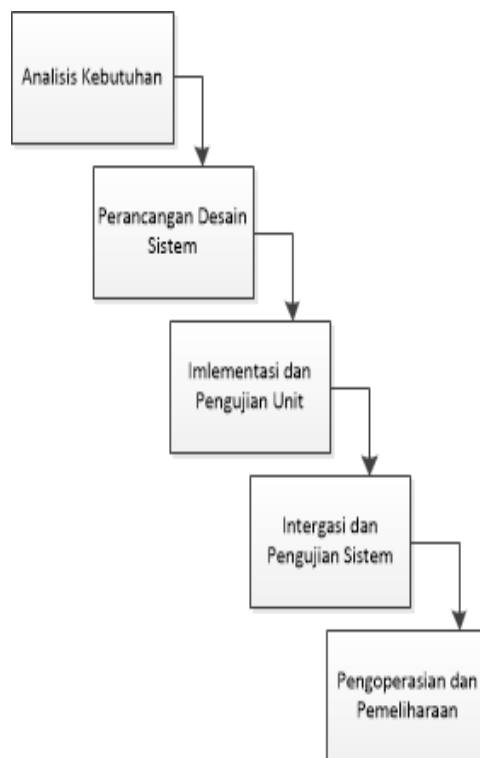




Gambar 1. Diagram Alur Penelitian

Perancangan Sistem

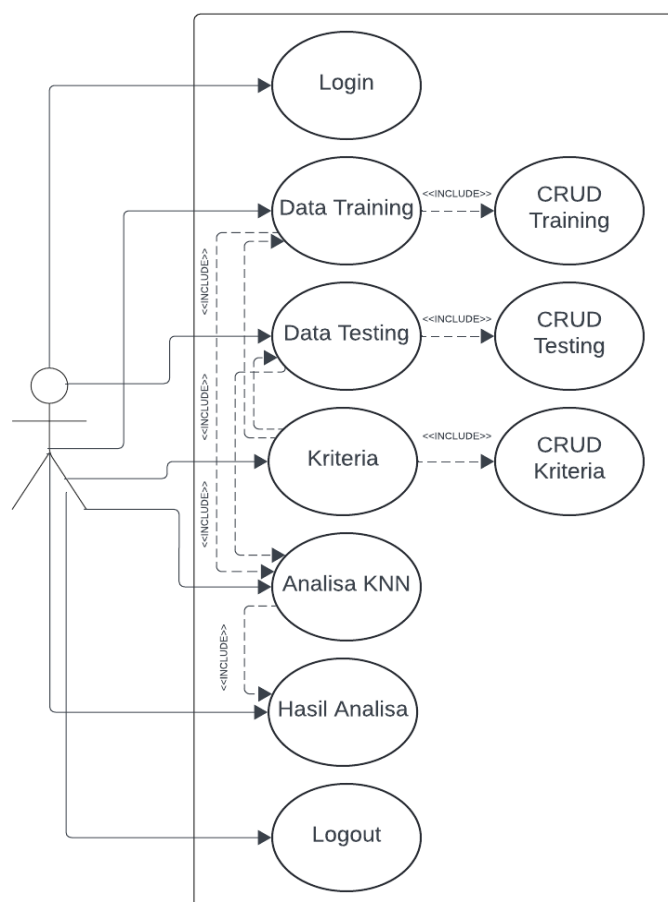
Perancangan perangkat lunak yang digunakan adalah menggunakan metode *Waterfall*. Metode *waterfall* adalah salah satu jenis model pengembangan perangkat lunak. Sebuah metode yang cara kerjanya seperti air terjun yang terdapat beberapa bagian metode dari awal hingga akhir.



Gambar 2. Perancangan Perangkat Lunak

Use Case Diagram

Alur *Use Case* pada penelitian ini yaitu admin memasukkan *username* dan *password* pada halaman login kemudian akan tampil beberapa menu *website*. Admin dapat mengakses menu yang ada di *website* seperti beranda, data *training*, data *testing*, kriteria, analisa dan hasil analisa. Pada data *training* dan data *testing* terdapat *CRUD* yang artinya kita bisa menambahkan aksi tambah, hapus, *update* dan *delete*. Pada menu analisa akan muncul proses perhitungan dan akan ditampilkan pada menu hasil analisa.



Gambar 3. Use Case Diagram

HASIL PENELITIAN DAN PEMBAHASAN

Data Awal

Pada penelitian ini menggunakan data sebanyak 182 data yang telah diperoleh dalam proses pengumpulan data sebelumnya. Data tersebut terdiri dari nama, IPA, IPS, Matematika, tes psikologi dan kategori yang bisa dilihat dibagian lampiran.

Preprocessing Data

Selanjutnya masuk ke dalam tahap *Preprocessing*. Pada tahap ini data akan di *filter* untuk menghilangkan atribut atau nilai yang kosong atau *missing* data sehingga data yang akan diolah menggunakan aplikasi *RapidMiner* sudah relevan atau sudah sesuai.

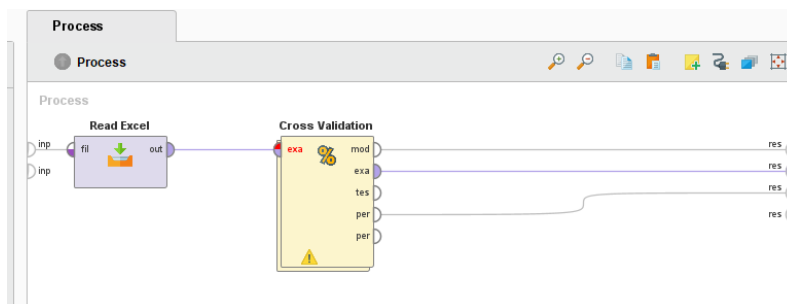
Tabel 1. Data Preprocessing

No	NAMA	IPA	IPS	MTK	TES PSIKOLOGI	KATEGORI
1	Aisyah Al Mumtahanah	80	79	80	113	IPA
2	Aisyah Raihana Robbani	79	79	77	107	IPS
3	Amelia Rahmah	81	82	81	111	IPA
...	...	...	...	...	...	...
180	Ryan Rahmawan	80	81	80	115	IPS
181	Siji Manggar Nasution	78	80	78	115	IPS
182	Abdullah Jafar Basalamah	78	79	77	111	IPS

Implementasi KNN

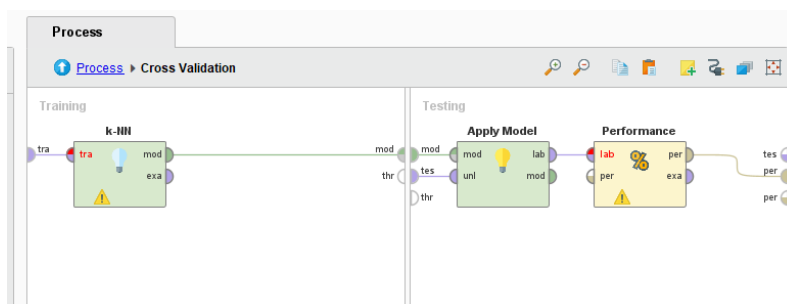
Pengujian *Cross Validation*

Hubungkan operator *read excel* yang digunakan untuk memasukkan data ke operator *Cross Validation* yaitu dengan ujung out pada read excel ke exa dan ujung mod, exa dan per pada *Cross Validation* disambungkan ke res.



Gambar 4. Rangkaian *Cross Validation* 1

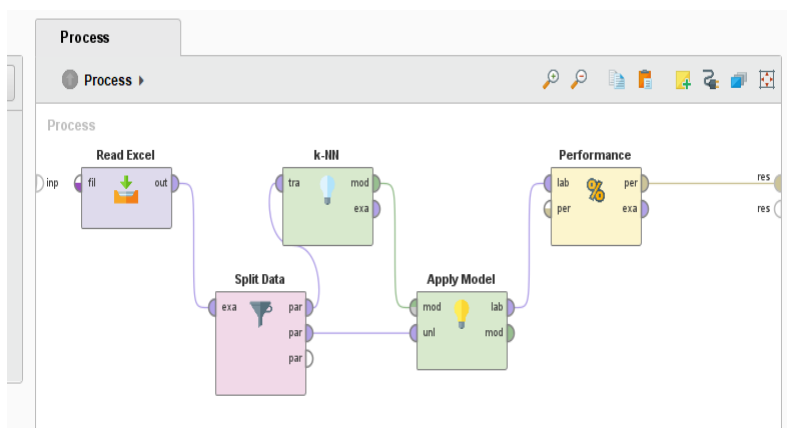
Selanjutnya double klik pada *Cross Validation* lalu rangkai operator *KNN* pada bagian *training* dengan ujung tra terhubung ke tra pada operator *KNN* dan mod ke mod pada *RapidMiner*. Kemudian pada bagian *testing*, hubungkan operator *Apply Model* ke operator *performance* dengan ujung mod ke mod *Apply Model* dan ujung lab ke lab operator *performance*.



Gambar 5. Rangkaian *Cross Validation* 2

Pengujian *Split Data*

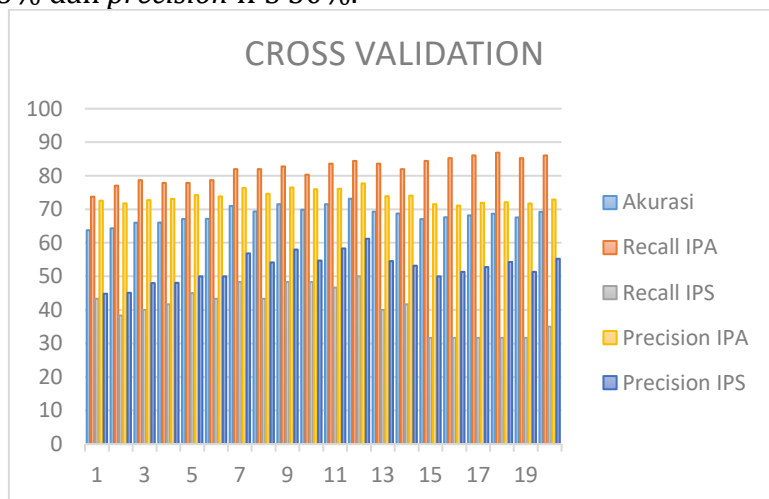
Pada pengujian *split data*, hubungkan operator *read excel* ke *split data* untuk proses pembagian data kemudian ujung par pada operator split data dihubungkan ke tra pada operator *KNN* dan unl pada operator *Apply Model*. Hubungkan juga ujung mod pada operator *KNN* ke mod pada operator *Apply model* dan ujung yang lain ke operator *performance*.



Gambar 6. Rangkaian *Split Data*

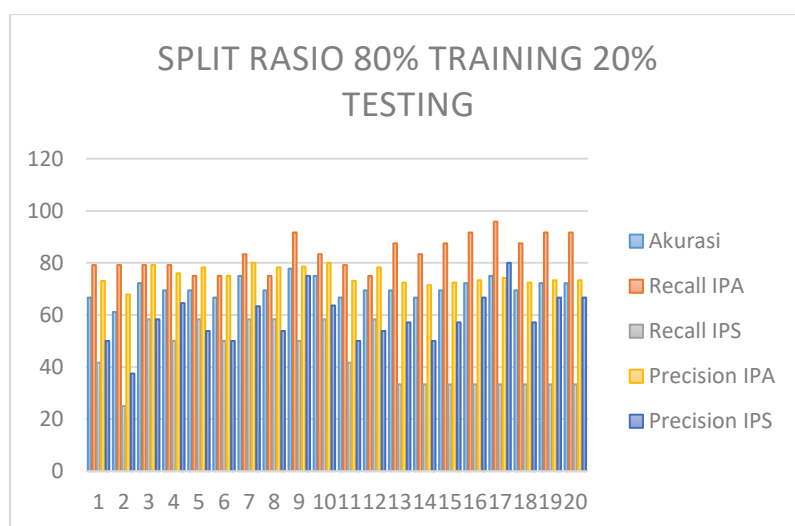
Pengujian Data

Pengujian *Cross Validation* dilakukan sebanyak 20 kali dengan nilai K=1 sampai dengan K=20 menghasilkan nilai akurasi tertinggi diperoleh pada K=12 dengan akurasi 73,16%, *class recall* IPA 84,43%, *class recall* IPS 50%, *precision* IPA 77,74% dan *precision* IPS 61,22% dan nilai terendah pada K=15 dengan akurasi 67,08%, *class recall* IPA 84,43%, *class recall* IPS 31,67%, *precision* IPA 71,53% dan *precision* IPS 50%.



Gambar 7. Uji Cross Validation

Pengujian selanjutnya yaitu dengan membagi 80% data *training* dan 20% data *testing*. Pengujian dilakukan sebanyak 20 kali dengan nilai K=1 sampai dengan K=20 menghasilkan nilai akurasi tertinggi diperoleh pada K=9 dengan akurasi sebanyak 77,78%, *class recall* IPA 91,67%, *class recall* IPS 50%, *precision* IPA 78,57% dan *precision* IPS 75% dan nilai terendah pada K=2 dengan akurasi 61,11%, *class recall* IPA 79,17%, *class recall* IPS 25%, *precision* IPA 67,86% dan *precision* IPS 37,50%



Gambar 8. Uji Split Rasio 80% Training 20% Testing

Evaluasi Hasil

Berdasarkan hasil beberapa pengujian yang sudah dilakukan menghasilkan pengujian *split rasio* lebih efektif pada 182 data yang diuji dibandingkan dengan pengujian *Cross Validation*. Hasil pengujian *split rasio* tertinggi dengan pembagian 80% data *training* dan 20% data *testing* diperoleh pada K=9 dengan akurasi sebanyak 77,78%, *class recall* IPA 91,67%, *class*

recall IPS 50%, *precision* IPA 78,57% dan *precision* IPS 75%. Dengan hasil tersebut artinya klasifikasi jurusan dapat menggunakan metode *K-Nearest Neighbor* karena menghasilkan nilai akurasi yang cukup baik

KESIMPULAN

Berdasarkan hasil penelitian dan pengujian data yang dilakukan di SMA Al Irsyad Tegal tentang penerapan algoritma *K-Nearest Neighbor* untuk klasifikasi jurusan siswa kelas XI dapat diambil beberapa kesimpulan sebagai berikut:

1. Berdasarkan dari beberapa pengujian yang dilakukan menggunakan *RapidMiner* menghasilkan nilai akurasi tertinggi diperoleh pada $K=9$ dengan akurasi sebanyak 77,78%, *class recall* IPA 91,67%, *class recall* IPS 50%, *precision* IPA 78,57% dan *precision* IPS 75%. dengan pembagian 80% data *training* dan 20% data *testing*.
2. Dengan hasil tersebut dapat diartikan metode *K-Nearest Neighbor* dapat digunakan untuk klasifikasi jurusan siswa kelas XI di SMA Al Irsyad karena memiliki tingkat akurasi yang cukup baik.
3. Dengan adanya sistem berbasis website dapat mempermudah tim BK untuk klasifikasi jurusan siswa sehingga lebih efektif dan efisien.

DAFTAR PUSTAKA

- A. B. Santoso, Pemrograman Web PHP Dasar Database MYSQLI dengan Bootstrap, 2022nd ed. Bandung: Widina Bhakti Persada Bandung.
- A. Maulida, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," vol. 1, no. 2, pp. 29–33, 2020.
- A. Muhaimin, M. Amin Hariyadi, and M. I. Imamudin, "Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor dan Kedisiplinan dengan Metode K-Nearest Neighbor," J. Ilmu Komput. dan Sist. Inf., vol. 7, no. 1, pp. 193–202, 2024, doi: 10.55338/jikoms.v7i1.2865.
- A. Muzaky, F. Arifianto, R. Hendrowati, and M. Darwis, "Menerapkan Metode Klasifikasi pada Data Uji Emisi Kendaraan di Jakarta dengan Menggunakan Jupyter Notebook," vol. 5, no. 2, 2024.
- A. Muzani, M. I. A. Sukri, S. N. Fauziah, A. Fatkhurohman, and D. Ariatmanto, "Data Mining Untuk Klasifikasi Produk Menggunakan Algoritma K-Nearest Neighbor Pada Toko Online," pp. 141–145.
- A. Qurotul and M. Reza, "Application Of The K-Nearest Neighbor Algorithm For Student Department Classification At 15 Pekanbaru State High School Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru," vol. 3, no. 1, pp. 39–45, 2023.
- H. P. P. Agung Noviantoroa, Amelia Belinda Silvianab, Risma Rahmalia Fitrianic, "Rancangan Dan Implementasi Aplikasi Sewa Lapangan Badminton Wilayah Depok Berbasis Web," vol. 1, no. 2, pp. 88–103, 2022.
- J. Saputra, Y. Sa, V. Yoga Pudya Ardhana, and M. Afriansyah, "RESOLUSI : Rekayasa Teknik Informatika dan Informasi Klasifikasi Kematangan Buah Alpukat Mentega Menggunakan Metode K-Nearest Neighbor Berdasarkan Warna Kulit Buah," Media Online, vol. 3, no. 5, pp. 347–354, 2023, [Online]. Available: <https://djournals.com/resolusi>
- M. Permadi, F. Febriani, and M. J. Panggabean, "Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Klasifikasi Penjualan Produk Indihome Daerah Riau," Semin. Nas. Penelit. dan Pengabd. Masy., pp. 105–111, 2023.
- N. Meilani and O. Nurdian, "Data Mining untuk Klasifikasi Penderita Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor," J. Wahana Inform., vol. 2, no. 1, pp. 177–187, 2023, [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer>.

- R. Sitepu, "Implementasi Algoritma K-Nearest Neighbor Untuk Klasifikasi Pengajuan Kredit," vol. 1, pp. 49–56, 2022.
- V. Wulandari, W. J. Sari, and Z. Alfian, "Implementation of Naïve Bayes Classifier and K-Nearest Neighbor Algorithms for Chronic Kidney Disease Classification Implementasi Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor untuk Klasifikasi Penyakit Ginjal Kronik," vol. 4, no. April, pp. 710–718, 2024.