



## Perbaikan *Missing Value* Menggunakan Pendekatan Korelasi Pada Metode *K-Nearest Neighbor*

Novta Dany'el Irawan<sup>1</sup>, Wijono<sup>2</sup>, Onny Setyawati<sup>3</sup>

<sup>1,2,3</sup> Jurusan Teknik Elektro, Universitas Brawijaya, Indonesia

<sup>1,2,3</sup> Ketawanggede, Kec. Lowokwaru, Kota Malang, Jawa Timur 65145, Indonesia

Email korespondensi: [novtadirawan@gmail.com](mailto:novtadirawan@gmail.com)

Dikirim 7 Agustus 2017, Direvisi 17 Agustus 2017, Diterima 20 Agustus 2017

Abstrak – *Missing value* sering terjadi dalam metode klasifikasi dikarenakan informasi tentang obyek tidak diberikan, sulit dicari atau memang informasi tersebut tidak ada. Hal ini menyebabkan menurunnya keakuratan dan kualitas data pada saat data diolah. Pendekatan korelasi dilakukan karena peneliti harus mengetahui tentang ada tidaknya dan kuat lemahnya hubungan *variable* yang terkait dalam suatu objek atau subjek yang diteliti. Metode klasifikasi yang digunakan yaitu metode K-NN, karena metode ini termasuk metode klasifikasi yang memiliki konsistensi yang kuat. Metode ini mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan nilai K atau tertangga terdekat. Pendekatan korelasi dapat dilakukan untuk mengatasi *missing value*, terbukti dengan meningkatnya hasil klasifikasi dan hilangnya data yang belum terklasifikasi. Kuisioner berfungsi sebagai alat ukur, kuisioner berisi beberapa pertanyaan yang diberikan kepada responden, dari hasil kuisioner dilakukan analisa data untuk mengetahui tingkat korelasi data cadangan. Setelah mendapatkan tingkat korelasi data cadangan, maka data cadangan tersebut digunakan sebagai pengganti data yang terdapat *missing value*. Sebelum dilakukan penggantian data yang terdapat *missing value*, hasil klasifikasi dari 500 data adalah jurusan IPA sejumlah 88 siswa, jurusan IPS 126 siswa, jurusan bahasa 271 siswa, dan belum terklasifikasi/*false* 15 siswa. Setelah dilakukan penggantian data yang terdapat *missing value*, hasil klasifikasi dari 500 data adalah jurusan IPA berjumlah 102 siswa, jurusan IPS berjumlah 316 siswa, bahasa berjumlah 82 siswa, dan tidak ada data yang belum terklasifikasi. Berdasarkan hasil percobaan yang dilakukan nilai  $k = 3, 5, 7, 9, \text{ dan } 11$ . Dapat diketahui nilai  $k = 5$  memiliki tingkat akurasi yang tinggi dibandingkan dengan nilai  $k$  yang lain yaitu 97%, jadi dalam penelitian ini nilai  $k$  yang dipakai pada metode K-NN adalah 5.

Kata kunci – korelasi, *missing value*, *K-nearest neighbor*

Abstract - *Missing value* often occur in classification method that is caused by information on the object is not given, it is difficult to find, or because of the information is unavailable. It will cause the decrement of accuracy and data quality during it is analyzed. Correlation approach was conducted because it should be known the existence and the strength of variable correlation in related to an object or subject studied. Classification method used is K-NN method. It is because this method is included in classification method that has strong consistency by finding the case through calculation on the closeness between the case with the old one based on K value or the nearest neighbor. Correlation approach can be done to overcome missing value, as evidenced by the increasing classification results and the loss of unclassified data. Questionnaire as a measuring tool, the questionnaire contains some questions given to the respondent, from the results of questionnaires conducted data analysis to determine the level of correlation of data backup. After getting the level of backup data correlation, then the backup data is used as a substitute for missing data value. Before the replacement of data there is missing value classification of 500 data classified natural science major 88 students, social science major 126 students, the language major 271 students, and unclassified / false 15 students. After the replacement of data there is missing value from 500 data, it can be classified into natural science major 102 students, social science major 316 students, the language major 82 students, and no unclassified data. Based on the experimental results, the value of  $k = 3, 5, 7, 9, \text{ and } 11$ . It can be seen that  $k = 5$  has a high accuracy of 97.0%, so in this study majors using K-NN method set  $k$  value used is 5.

Keywords – correlation, missing value, K-nearest neighbor

## I. PENDAHULUAN

Terjadinya *missing value* disebabkan karena informasi tentang objek tidak diberikan, sulit dicari, atau memang informasi tersebut tidak ada, hal ini akan menyebabkan menurunnya keakuratan dan kualitas data pada saat data diolah. *Missing value* juga sering terjadi karena responden tidak memberikan jawaban pada alternatif jawaban yang disediakan [1]. Seperti pada kasus penjurusan siswa, *missing value* sering terjadi karena siswa dan orang tua sering mengabaikan pentingnya pengisian data yang nanti akan berfungsi sebagai bahan pertimbangan dalam penentuan hasil penjurusan.

Prosedur operasional rutin pada Sekolah Menengah Atas (SMA) adalah memberi kesempatan kepada para siswa untuk memilih jurusan yang cocok [2]. Siswa yang akan mengikuti penjurusan harus memberikan data pendukung untuk diproses dalam sistem penjurusan, data pendukung tersebut kemudian diproses berdasarkan beberapa kriteria. Dalam mempercepat proses klasifikasi penjurusan siswa SMA diperlukan sebuah Sistem Pendukung Keputusan (SPK) yang dapat membantu sekolah untuk mengambil keputusan yang cepat, dan tepat [3]. Seperti pada penelitian sebelumnya, untuk hasil yang cepat dan tepat dalam penjurusan siswa SMA maka dibangun sebuah SPK dengan menggunakan metode berbasis komputer *Profile Matching* [4]. Sebuah SPK akan berfungsi jika data pendukungnya lengkap, tapi fakta di lapangan *user* sering tidak melengkapi data pendukung tersebut, sehingga data pendukung menjadi kosong (*missing value*), hal ini menyebabkan hasil penjurusan tidak selalu sama dengan keinginan serta harapan dari siswa atau orang tua.

Mengganti *variable* kosong dengan estimasi dari beberapa nilai tidak dapat langsung dilakukan jika *variable* kosong berupa data kualitatif, perlu ada pendekatan khusus untuk mencari hubungan kedekatan *variable* kosong (*missing value*) dengan *variable* pengganti atau cadangan yang juga berupa data kualitatif [5]. Pendekatan yang dapat digunakan untuk mencari kedekatan antara *variable* kosong dengan *variable* cadangan yaitu pendekatan korelasi. Pendekatan korelasi dilakukan karena peneliti harus mengetahui tentang ada tidaknya dan kuat lemahnya hubungan *variable* yang terkait dalam suatu objek atau subjek yang diteliti. Adanya hubungan dan tingkat *variable* ini menjadi penting, karena dengan mengetahui tingkat hubungan yang ada, peneliti akan dapat mengembangkannya sesuai dengan tujuan penelitian. Dalam ilmu statistika istilah korelasi diberi pengertian sebagai hubungan antara dua *variable* atau lebih [6]. Pendekatan korelasi sendiri bertujuan untuk mencari bukti terdapat tidaknya hubungan (korelasi) antar *variable*, selanjutnya jika sudah ada hubungan, tujuannya yaitu untuk melihat tingkat kedekatan hubungan antar *variable*, dan untuk memperoleh kejelasan dan kepastian apakah hubungan tersebut meyakinkan atau tidak meyakinkan [7].

Sejak diperkenalkannya metode *machine learning* untuk proses klasifikasi, banyak algoritma klasifikasi yang mulai diaplikasikan dalam SPK, diantaranya yaitu metode *K-Means*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbor* (K-NN), tetapi untuk metode *K-means* dan SVM mempunyai kelemahan yaitu sulit dipakai dalam problem berskala besar, skala besar dalam hal ini dimaksudkan dengan jumlah sampel yang diolah [8]. Berbeda dengan metode K-NN, metode ini termasuk metode klasifikasi data yang memiliki konsistensi yang kuat, dengan cara mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan bobot [9].

Pada penelitian lain diungkapkan dalam masalah klasifikasi dan sistem pendukung keputusan dengan menerapkan metode K-NN pada pendukung keputusan beasiswa, algoritma ini lebih efektif dalam melakukan *training* data yang besar dan dapat menghasilkan data yang lebih akurat [10]. Berdasarkan beberapa penelitian terdahulu, dalam penelitian ini alternatif solusi yang bisa diberikan dalam mengatasi data yang mengandung *missing value* dengan cara menambahkan *variable* data cadangan yang diperoleh dari hasil pengukuran menggunakan pendekatan korelasi dengan metode klasifikasi K-NN. Dengan diatasinya *missing value* diharapkan proses penjurusan menjadi lebih tepat, cepat dan efisien.

## II. METODE PENELITIAN

### A. Persiapan Data

Alat ukur dengan cara diberikan kuisioner dengan beberapa pertanyaan kepada responden. Pembuatan kuisioner ini mengacu parameter yang sudah dibuat oleh peneliti terhadap penelitian yang dilakukan. Pertanyaan pada bagian pertama digunakan untuk mengetahui identitas responden, bagian kedua untuk mengetahui data *variable* pendukung pribadi responden yang diprediksi mempunyai hubungan dengan penjurusan, bagian ketiga tingkat akurasi hasil penjurusan manual yang dilakukan selama ini. Apabila ada responden yang mengalami kesulitan dalam membaca kuisioner, maka peneliti akan membantu membacakan. Setelah kuisioner sebagai alat penelitian selesai disusun kemudian dilakukan uji validitas oleh ahli.

Jenis data dan cara pengumpulan data:

- a) Data *Training*: Data kasus lama, diambil dengan cara meng-input pola kasus.
- b) Data *Testing*: Data siswa baru, diambil dengan cara dokumentasi.
- c) Data *variable* inti:
  1. Data *variable* 1: Nilai Report, diambil dengan cara dokumentasi.
  2. Data *variable* 2: Nilai Psikotes, diambil dengan cara isian hasil tes.
  3. Data *variable* 3: Bakat dan Minat, diambil dengan cara wawancara oleh guru BK.

4. Data *variable* 4: Pendapatan Orang Tua, diambil dengan cara isian data oleh orang tua
- d) Data *variable* cadangan:
  1. Data *variable* Hobi, diambil dengan cara isian data.
  2. Data *variable* Cita – cita, diambil dengan cara isian data.
  3. Data *variable* Pekerjaan Ortu diambil dengan cara isian data.
- e) *Variable* tambahan: Data *variable* hasil tes kesehatan, diambil dengan cara isian data hasil tes

Dari hasil uji validasi ahli untuk klasifikasi penjurusan dihasilkan beberapa cadangan untuk menggantikan data inti, tetapi data cadangan tersebut tidak bisa langsung dipakai untuk menggantikan data inti. Diperlukan analisa data seberapa kuat tingkat korelasinya dengan data inti.

### B. Analisis Data

Metode analisis ini digunakan untuk memberikan gambaran fenomena penelitian yaitu tentang: 1. hubungan data bakat minat dengan data cadangan hobi dan cita – cita; 2. hubungan data pendapat orang tua dengan data cadangan pekerjaan orang tua dan tes kesehatan (buta warna). Proses pendekatan korelasi dapat dilihat pada Gambar 1. Skoring dapat dilakukan untuk mengetahui seberapa kuat hubungan antara data, proses skoring diperoleh dari persamaan (1).

$$P = \frac{n}{N} \times 100\% \quad (1)$$

Dimana :

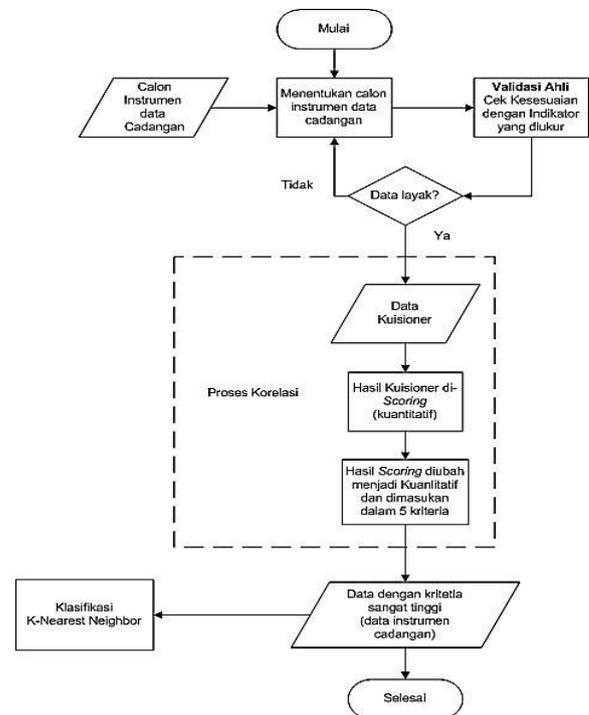
- P : Persentase nilai yang diperoleh  
 n : Jumlah skor yang diperoleh  
 N : Jumlah seluruh nilai ideal, dicari dengan cara jumlah item dikalikan nilai ideal tiap item.

Hasil kuantitatif dari perhitungan dengan persamaan selanjutnya diubah atau ditafsirkan dengan kalimat yang bersifat kualitatif. Dalam hal ini *variable* data cadangan pada penjurusan siswa ditafsirkan secara kualitatif ke dalam lima kriteria. Tahapan untuk menentukan jenjang kriteria tersebut dilakukan dengan cara berikut :

- a) Presentase maksimal dapat ditetapkan dari hasil persamaan (2).
 
$$(4:4) \times 100\% = 100\% \quad (2)$$
- b) Untuk menetapkan persentase minimal menggunakan persamaan (3).
 
$$(1:4) \times 100\% = 25\% \quad (3)$$
- c) Menetapkan rentang persentase yang diperoleh dengan cara mengurangi persentase maksimal (100%) dengan persentase minimal (25%) yaitu 75%.
- d) Menetapkan panjang kelas interval persentase yang diperoleh dengan cara membagi rentang

persentase dengan banyaknya kirteria. Banyaknya kriteria yang dipakai adalah sejumlah lima kriteria yakni sangat tinggi, tinggi, sedang, rendah sangat rendah sehingga panjang kelas interval persentasenya yaitu  $75\% : 5 = 15\%$ .

- e) Menetapkan jenjang kriteria, berdasarkan perhitungan di atas, maka jenjang kriteria tingkat korelasi data *missing value* dengan data cadangan dapat dilihat pada tabel 1.



Gambar 1. Flowchart Pendekatan Korelasi

Tabel 1. Kriteria Tingkat Korelasi

Interval	Kriteria
86% - 100%	Sangat Tinggi
71% - 85%	Tinggi
56% - 70%	Sedang
41% - 55%	Rendah
25% - 40%	Sangat Rendah

Kriteria penilaian hubungan data *missing value* dengan data cadangan diatas akan mempermudah peneliti dalam menentukan gambaran hubungan antara data *missing value* dengan data cadangan pada penjurusan siswa SMA menggunakan metode K-NN.

### C. Klasifikasi K-Nearest Neighbor

Data awal akan dibagi menjadi dua bagian yaitu data *training* (latih) dan data *testing* (uji). Selanjutnya akan dilakukan perhitungan terhadap data uji dengan menggunakan metode K-Nearest Neighbor untuk menghasilkan probabilitas setiap kelas dengan menghitung bobot kedekatannya dengan data latih yang ada. Nilai probabilitas setiap kelas yang dihasilkan, digunakan untuk melakukan klasifikasi

data uji sehingga menghasilkan data uji yang terklasifikasi.

Tidak ada metode khusus, dalam penentuan nilai k pada metode K-NN. Jika nilai k terlalu kecil maka akan banyak noise yang mengurangi tingkat akurasi dalam klasifikasi, tetapi jika terlalu besar juga dapat menyebabkan kesalahan dalam pembatasan nilai yang diambil dan secara tidak langsung juga mempengaruhi akurasinya.

#### D. Uji Akurasi Hasil Klasifikasi

Dalam penelitian ini, uji akurasi dilakukan dengan cara membandingkan secara manual data terklasifikasi benar dan data terklasifikasi salah dari data *training* dengan data *testing*. Data *training* adalah data yang telah diklasifikasi secara manual sebelumnya dan tidak terdapat *missing value*, data *testing* adalah data yang mengandung *missing value* dan sudah dilakukan perbaikan dengan pendekatan korelasi dan diklasifikasi dengan metode K-NN.

### III. HASIL PENELITIAN

#### A. Hasil Tingkat Korelasi Instrumen Cadangan

Setelah *variable* data cadangan divalidasi dan kuisioner disusun oleh ahli, selanjutnya dilakukan pengambilan data dengan penyebaran kuisioner terhadap 500 responden. Hasil data yang diperoleh dari kuisioner selanjutnya diskoring dan dimasukkan dalam lima kriteria sesuai dengan Tabel 1.

Berdasarkan Tabel 2 diketahui bahwa tingkat korelasi *variable* cadangan hobi dengan kriteria sangat tinggi berjumlah 32 responden (6,4%), kriteria tinggi berjumlah 462 responden (92,4%) dan kriteria sedang berjumlah 6 responden (1,2%). Sedangkan untuk *variable* cadangan cita – cita dengan kriteria sangat tinggi berjumlah 9 responden (1,8%), kriteria tinggi berjumlah 435 responden (87%), dan kriteria sedang 56 responden (11,2%). Dari hasil skoring data kuisioner tersebut diketahui tingkat korelasi *variable* cadangan hobi dan *variable* cadangan cita- cita adalah layak untuk digunakan sebagai pengganti *variable* inti yang mengalami *missing value*.

Tabel 2. Tingkat Korelasi Variable Cadangan Hobi dan Cita-Cita

Kriteria	Variable cadangan Hobi		Variable cadangan Cita - cita	
	Jumlah	Persentase	Jumlah	Persentase
Sangat Tinggi	32	6,4%	9	1,8%
Tinggi	462	92,4%	435	87%
Sedang	6	1,2%	56	11,2%
Rendah	0	0%	0	0%
Sangat Rendah	0	0%	0	0%

Tingkat korelasi *variable* cadangan pekerjaan ortu dan *variable* tambahan tes kesehatan dapat dilihat pada Tabel 3 diketahui bahwa korelasi *variable* cadangan pekerjaan ortu dengan kriteria tinggi berjumlah 55 responden (11%), kriteria sedang

berjumlah 433 responden (86,6%) dan kriteria rendah berjumlah 12 responden (2,4%). Sedangkan untuk *variable* tambahan tes kesehatan dengan kriteria tinggi berjumlah 201 responden (40,2%), kriteria sedang berjumlah 298 responden (59,6%), dan rendah 1 responden (0,2%). Dari hasil skoring data kuisioner tersebut diketahui tingkat korelasi *variable* cadangan pekerjaan dan *variable* tambahan tes kesehatan adalah masih layak untuk digunakan sebagai pengganti *variable* inti yang mengalami *missing value*.

Tabel 3. Tingkat Korelasi Variable Cadangan Pekerjaan Orang Tua dan Tes Kesehatan

Kriteria	Variable cadangan Pekerjaan ortu		Variable cadangan Tes kesehatan	
	Jumlah	Persentase	Jumlah	Persentase
Sangat Tinggi	0	0%	0	0%
Tinggi	55	11%	201	40,2%
Sedang	433	86,6%	298	59,6%
Rendah	12	2,4%	1	0,2%
Sangat Rendah	0	0%	0	0%

#### B. Pola Kasus

Pola kasus ini berfungsi sebagai bahan uji terhadap data yang mengalami *missing value*, pola kasus tersebut dapat juga disebut data *training*. Pada Tabel 4 terdapat pola kasus *variable* yang tidak mengalami *missing value*, yang nanti kasus tersebut digunakan sebagai data *training*, selanjutnya dilakukan uji coba dengan menerapkan *missing value* pada beberapa *variable* data *testing* yang sudah mempunyai *variable* cadangan dari hasil korelasi.

Tabel 4. Pola Kasus Tidak Terdapat Missing Value

No	ID Kasus	V 1	V 2	V 3	V 4	Jurusan
1	case 1	1	1	1	1	Bahasa
2	case 2	2	1	1	1	Bahasa
3	case 3	1	2	1	1	Bahasa
4	case 4	1	1	2	1	Bahasa
5	case 5	1	1	1	2	Bahasa
6	case 6	3	1	1	1	Bahasa
7	case 7	1	3	1	1	Bahasa
8	case 8	1	1	3	1	Bahasa
9	case 9	1	1	1	3	Bahasa
10	case 10	2	2	2	2	IPS
11	case 11	2	1	2	2	IPS
12	case 12	2	2	1	2	IPS
13	case 13	2	2	2	1	IPS
14	case 14	2	3	2	2	IPS
15	case 15	2	2	3	2	IPS
16	case 16	2	2	2	3	IPS
17	case 17	3	3	3	3	IPA
18	case 18	3	1	3	3	IPA
19	case 19	3	3	1	3	IPA
20	case 20	3	3	3	1	IPA
21	case 21	3	2	3	3	IPA
22	case 22	3	3	2	3	IPA
23	case 23	3	3	3	2	IPA
24	case 24	2	3	1	1	IPS
25	case 25	1	2	3	1	IPS

IV. PEMBAHASAN

A. Penggantian Data Missing Value dengan Data Cadangan

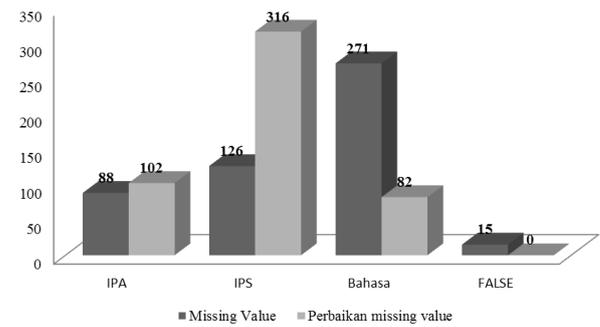
Setelah mendapatkan tingkat korelasi data cadangan maka data cadangan tersebut dapat digunakan pengganti data yang terdapat *missing value*, data yang terdapat *missing value* di asumsikan pada *variable 3* dan *variable 4*, lihat pada Tabel 5. Dalam penelitian ini data cadangan diuji dalam beberapa kasus, kasus tersebut dapat juga disebut dengan data *training*. Data *training* dapat dilihat pada Tabel 5.

Tabel 5. Contoh Data Terdapat Missing Value

No	ID Data	V1	V2	V3	V4	Jurusan
1	Missing value 1	3	2		1	Bahasa
2	Missing value 2	2	3			Bahasa
3	Missing value 3	2	3	3		IPA
4	Missing value 4	1	1		2	Bahasa
5	Missing value 5	2	3		3	IPA
6	Missing value 6	3	2	3		IPA
7	Missing value 7	2	1		3	Bahasa
8	Missing value 8	2	1	2		Bahasa
9	Missing value 9	2	3	3		IPA
10	Missing value 10	2	1	1		Bahasa
11	Missing value 11	3	3	3		IPA
12	Missing value 12	1	3			Bahasa
13	Missing value 13	3	3	1	3	IPA
14	Missing value 14	1	3	1		Bahasa
15	Missing value 15	2	3	3		IPA
16	Missing value 16	2	2		1	Bahasa
17	Missing value 17	3	1	1		Bahasa
18	Missing value 18	1	3	3		IPS
19	Missing value 19	3	2	3		IPA
20	Missing value 20	1	1		1	FALSE

Dapat diketahui dari Tabel 5. masih ada data yang belum terklasifikasi, hal ini disebabkan data tersebut oleh K-NN dibaca berada diluar kriteria penjurusan, salah satu faktornya yaitu masih terdapat *missing value* yang ditandai dengan arsiran warna hitam pada tabel 5.5. Dari 500 data yang masih terdapat *missing value* terklasifikasi jurusan IPA sejumlah 88 siswa, jurusan IPS 126, jurusan bahasa 271, dan belum terklasifikasi/false 15.

Dari 500 data *missing value* yang sudah digantikan dengan data cadangan, diperoleh data yang dapat terklasifikasi sebagai berikut: jurusan IPA berjumlah 102 siswa, jurusan IPS berjumlah 316 siswa, dan bahasa berjumlah 82 siswa. Sebagian data dapat dilihat pada Tabel 6. Untuk perbandingan lebih jelas antara data *missing value* dan data yang dilakukan perbaikan *missing value* dapat dilihat pada Gambar 2.



Gambar 2. Grafik Perbandingan Hasil Klasifikasi Sebelum dan Sesudah Perbaikan

Tabel 6. Data Missing Value Diganti Variable Cadangan dan Terklasifikasi

No	ID Data	V1	V2	V3	V3''	V4	V4''	Jurusan
1	Missing value 1	3	2		2	1		IPS
2	Missing value 2	2	3		2		3	IPA
3	Missing value 3	1	1	2			3	IPS
4	Missing value 4	1	1		3	2		IPS
5	Missing value 5	2	3		2	3		IPA
6	Missing value 6	3	3	2			2	IPA
7	Missing value 7	2	1		1	3		IPS
8	Missing value 8	2	1	2			2	IPS
9	Missing value 9	2	3	3			1	IPS
10	Missing value 10	2	1	1			1	Bahasa
11	Missing value 11	3	3	3			1	IPA
12	Missing value 12	1	3		1		3	IPS
13	Missing value 13	3	3		3	3		IPA
14	Missing value 14	1	3	1			3	IPS
15	Missing value 15	2	3	3			1	IPA
16	Missing value 16	2	2		3	1		IPS
17	Missing value 17	3	1	1			1	Bahasa
18	Missing value 18	1	3	3			2	IPS
19	Missing value 19	3	2	3			1	IPS
20	Missing value 20	1	1		2	1		Bahasa

### B. Pengujian Akurasi

Tahap akhir dalam penelitian ini adalah menentukan akurasi metode K-NN terhadap data testing yang telah diklasifikasikan. Untuk mendapat tingkat akurasi baik maka dilakukan percobaan nilai  $k = 3, 5, 7, 9$ , dan  $11$  kemudian dibandingkan hasilnya untuk diketahui tingkat akurasi yang lebih baik.

Tabel 7. Tabel Perbandingan dan Tingkat Akurasi Nilai K

nilai k	Klasifikasi Benar	Klasifikasi Salah	Akurasi
3	470	30	94,0%
5	485	15	97,0%
7	483	18	96,6%
9	478	22	95,6%
11	473	27	94,6%

Hasil percobaan nilai  $k$  pada Tabel 7. dapat diketahui dari 500 data penjurusan dan pola kasus sesuai Tabel 4, maka diperoleh nilai akurasi yang hampir sama. Nilai  $k = 5$  memiliki tingkat akurasi yang tinggi yaitu 97,0% , jadi dalam penelitian ini

penjurusan menggunakan metode K-NN ditetapkan nilai  $k$  yang dipakai adalah 5.

### C. Proses Klasifikasi k-NN

Proses klasifikasi dengan metode K-NN menggunakan nilai  $k = 5$  terhadap data *missing value* yang sudah di ganti dengan data cadangan, dapat dilihat pada Tabel 8.

Berdasarkan Tabel 8, jika nilai  $k$  yang digunakan 5 maka akan didapatkan hasil Jurusan IPA = 0, IPS = 3, Bahasa = 2, dari lima pola kasus IPS memiliki jumlah paling banyak jadi kesimpulannya adalah siswa yang memiliki pola kasus sesuai data *testing* telah diprediksi masuk kedalam jurusan IPS.

Pola setiap *variable* pada kasus Tabel 4 atau data *training* akan direkam, hal ini berfungsi jika ada data baru atau data *testing* metode K-NN akan bekerja dengan mengikuti pola terdekat pada data *training* tersebut, nilai  $k$  yang dipakai dalam penelitian ini adalah 5, jadi jika ada data baru maka sistem akan mencari 5 pola *variable* kasus lama terdekat dengan pola kasus baru tersebut.

Tabel 8. Percobaan Nilai K

Nilai k	Kasus lama yang tidak terdapat <i>missing value</i>					Kasus baru <i>variable</i> sudah diganti <i>variable</i> cadangan				
	V1	V2	V3	V4	Jurusan	V1	V2	V3	V4'	Jurusan
5	1	3	1	1	Bahasa					
	1	1	3	1	Bahasa					
	1	2	1	3	IPS	1	1	3	2	?
	1	1	2	3	IPS					
	1	2	3	1	IPS					

## V. PENUTUP

### A. Kesimpulan

Dari peneitian ini dapat diambil beberapa kesimpulan yaitu pendekatan korelasi dapat dilakukan untuk mengatasi *missing value* yang terdapat pada SPK untuk penjurusan SMA. Metode K-NN digunakan sebagai metode klasifikasi dalam penelitian ini. Hasil percobaan nilai  $k = 3, 5, 7, 9$ , dan  $11$  dapat diketahui dari 500 data penjurusan dan 31 pola kasus lama, nilai  $k = 5$  memiliki tingkat akurasi yang tinggi yaitu 97,0% , jadi dalam penelitian ini penjurusan menggunakan metode K-NN ditetapkan nilai  $k$  yang dipakai adalah 5.

### B. Saran

Sebaiknya dilakukan penambahan variasi pola kasus lama. Sesuai dengan sifat yang dimiliki metode K-NN yaitu semakin banyak pola kasus maka akurasi klasifikasi akan bertambah. Penelitian ini dilakukan pada penjurusan siswa SMA, kasus *missing value* sering terjadi pada *variable* bakat dan minat, dan *variable* pendapat orang tua. Pendekatan korelasi dilakukan pada *variable* hobi dan *variable* pendapat orangtua sebagai data cadangan. Sebaiknya dilakukan

penelitian lebih lanjut untuk pendekatan korelasi pada *variable* selain *variable* hobi dan *variable* pekerjaan orang tua.

## DAFTAR PUSTAKA

- [1] T. Hendrawati, "KAJIAN METODE IMPUTASI DALAM MENANGANI MISSING DATA," *Pros. Semin. Nas. Mat. dan Pendidik. Mat. UMS*, pp. 637–642, 2015.
- [2] K. Pendidikan *et al.*, "Pedoman peminatan peserta didik," 2013.
- [3] A. A. Hartono, "Sistem Pendukung Keputusan pada Penjurusan Siswa Terkendala dengan metode Analytic Hierarchy Process," *JNTETI*, vol. 3, no. 3, pp. 194–200, 2014.
- [4] I. N. Farida and R. Firliana, "Implementasi Metode Profile Matching Untuk Evaluasi Potensi Akademik Penjurusan Siswa MAN 2 Kota Kediri," *Infotel Vol.8 No.2*, pp. 156–163, 2016.
- [5] I. Atastina, "Analisis penanganan *missing value* dengan metode collateral *missing value* estimation (cmve) analysis of *missing value* handling by collateral *missing value* estimation (cmve) method," 2011.
- [6] J. W. Creswell, *Educational Research*. 2012.
- [7] A. Sya, "Teknik analisis data penelitian," 2005.
- [8] A. Sulistiyo, J. T. Informatika, and F. Udinus, "Penentuan Jurusan Sekolah Menengah Atas

- Menggunakan Metode K-Nearest Neighbor Classifier Pada SMAN 16 Semarang,” pp. 1–5.
- [9] W. Kaswidjanti, N. Charibaldi, and D. L. Mallisa, “Pengembangan aplikasi pendukung keputusan untuk menentukan pekerjaan di bidang teknologi informasi,” *semnasIF 2010, UPN “Veteran” Yogyakarta*, vol. 2010, no. semnasIF, pp. 71–79, 2010.
- [10] Sumarlin, “Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM,” *ejournal.undip.ac.id*, vol. 1, pp. 52–62, 2015.