

Imputasi Missing Data Menggunakan Metode K-Nearest Neighbour Dengan Optimasi Algoritma Memetika

Missing Value Imputation Using KNN Method Optimized With Memetic Algorithm

Ida Bagus Gde Narinda Giriputra

Prodi S1 Ilmu Komputasi, Fakultas Informatika, Universitas Telkom

narinda.giriputra@gmail.com

Abstrak

Dalam suatu penelitian, data adalah hal yang sangat penting. Kualitas hasil dari penelitian akan berbanding lurus dengan kualitas data yang akan dipakai dalam penelitian yang bersangkutan. Salah satu permasalahan yang ada pada data set adalah ketiadaan nilai pada data untuk atribut tertentu atau yang lebih dikenal dengan istilah *missing data*. Salah satu metode yang sering digunakan oleh para peneliti adalah *k-Nearest Neighbour (KNN)*. Namun, metode ini memiliki beberapa kelemahan, salah satunya adalah pemilihan nilai *k* yang tidak tepat dapat menurunkan kinerja klasifikasi. Penelitian bertujuan menangani *missing data* dengan teknik *imputasi* menggunakan gabungan algoritma *KNN* dengan algoritma memetika. Algoritma memetika digunakan untuk mengoptimasi nilai *k* pada *KNN* sehingga dapat menghasilkan nilai estimasi yang baik dengan *mean square error (MSE)* sekecil mungkin. Pengujian performansi dilakukan dengan membandingkan nilai *MSE* hasil *imputasi missing data* pada saat *missing data* ada sebanyak, 5%, 10%, 15%, 20% dan 25% dengan menggunakan metode *KNN* dan *KNN* yang dioptimasi dengan algoritma memetika. Hasil dari penelitian menunjukkan bahwa *k* yang didapat dari hasil optimasi algoritma memetika tidak selalu menunjukkan *k* optimal dengan hasil pengukuran akurasi terbaik yang diuji, namun sangat mendekati hasil terbaik dari pengukuran akurasi yang diuji.

Kata Kunci : *missing data, KNN, algoritma memetika, optimasi*

Abstract

Data is one most important key in researching. The quality in one research will be highly affected by the quality of the data that researcher collect. One most deadly problem is missing data in the data set. One of many way to find missing data is k-Nearest Neighbour(KNN). But this method have a weakness. That is, k value that researcher don't know what number to set it. This project goal is to find a way to search optimal k value to input on KNN. Memetic algorithm is the metod that used in this journal project to find that. This project try to find optimal k value on KNN with memetic algorithm, that can minimize mean square error (MSE) as mush as possible. Performance testing is done by comparing the value of MSE from the imputation of missing data when there were missing data, 5%, 10%, 15%, 20% and 25% with KNN method and KNN with memethic algorithm optimization . Results from the study showed that the k obtained from the memetic algorithm optimization does not necessarily indicate the optimum k with the best accuracy measurement results are tested, but very close to the best results of the measurement accuracy is tested.

Keywords : *missing data, KNN, memetic algorithm, optimization*

1 Pendahuluan

Pengumpulan data merupakan bagian dari rangkaian riset yang pasti dilakukan oleh seluruh peneliti di dunia. Data merupakan bagian yang sangat penting dalam melakukan suatu penelitian. Hilangnya beberapa bagian data dalam suatu *data set* dapat memengaruhi ketepatan hasil riset yang sedang diuji, fenomena ini dikenal dengan istilah *missing data*. Fenomena *missing data* ini sangat sering terjadi dalam banyak riset. *Missing data* dapat terjadi karena beberapa hal, diantaranya kesalahan peneliti yang mengumpulkan data, alat

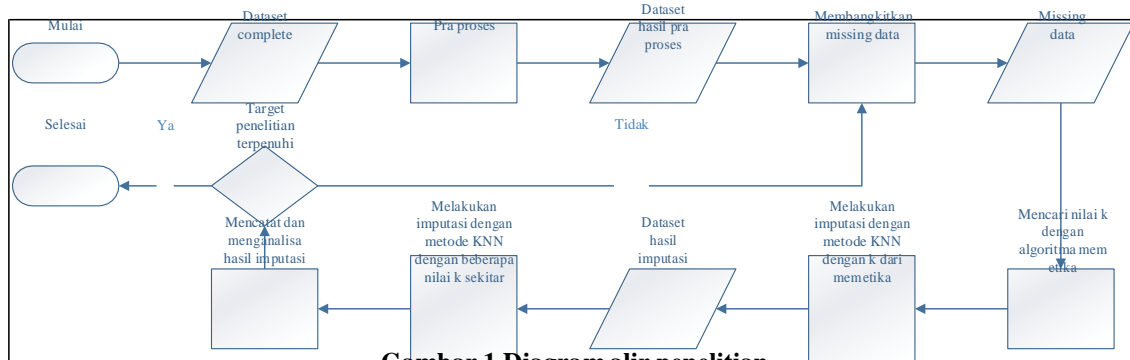
untuk mengumpulkan data yang terbatas, terjadinya kesalahan program yang mengumpulkan data, dan lain sebagainya. Pengaruh kurangnya data dalam suatu riset terlihat dalam perbedaan hasil penghitungan cepat dalam proses pemilihan presiden tahun 2014. Karena perbedaan data sampel yang diambil masing-masing lembaga riset mengakibatkan hasil penghitungan cepat yang berbeda juga, sehingga menyebabkan sedikit kegelisahan dalam masyarakat. Riset dalam berbagai bidang lainnya juga membutuhkan data yang lengkap demi ketepatan hasil riset, terutama riset dalam bidang yang berkaitan dengan keselamatan satu atau banyak orang. Maka dari itu dibutuhkan suatu alat untuk mengisi *missing data* dalam *data set* yang tidak sempurna.

Imputasi missing data adalah sebutan metode untuk mengisi *missing data* yang ada, beberapa diantaranya adalah menghilangkan nilai yang bersangkutan dengan data yang hilang tersebut, mengisi data yang hilang tersebut dengan perkiraan dari data yang memiliki kemiripan, dan metode lainnya. Para peneliti telah banyak memiliki cara untuk mengatasi *missing data* tersebut. Tugas akhir ini mencoba mengatasi *missing data* ini dengan cara *imputasi* melalui algoritma KNN.

Mencari nilai *missing data* dengan cara *imputasi KNN* bukanlah hal yang baru. Metode ini memiliki kelemahan dimana peneliti yang ingin mencari *missing data* dengan metode ini harus memasukkan nilai k , yang mana apabila nilai k bila diinput terlalu besar maupun terlalu kecil maka nilai *missing data* yang didapatkan akan berbeda jauh dari nilai yang semestinya ada. Oleh karena itu dalam tugas akhir ini akan mencari nilai k optimal dengan algoritma memetika. Algoritma memetika akan mencari nilai k yang optimal, sehingga peneliti mendapatkan nilai *imputasi missing data* dengan nilai kesalahan paling kecil. Diharapkan dari hasil tugas akhir ini, akan didapatkan suatu metode untuk melakukan *imputasi* pada *missing data* dengan kesalahan yang minimal untuk membantu para peneliti yang melakukan riset.

2 Metodologi

Penelitian ini menggunakan dataset lengkap yang digunakan untuk mengukur performansi algoritma memetika dalam mencari nilai k untuk melakukan imputasi missing data dengan metode KNN. Penelitian ini akan dilakukan dengan cara dibangkitkan missing data dari dataset komplit tersebut dengan proporsi 5%, 10%, 15%, 20%, dan 25% dari keseluruhan data. Langkah berikutnya adalah mencari nilai k optimal dengan algoritma memetika dalam melaksanakan imputasi terhadap missing data dengan menggunakan metode KNN. Selanjutnya k yang di dapat dari algoritma memetika akan diukur tingkat kekeliruannya dari nilai MSE, semakin kecil nilai MSE maka semakin bagus nilai imputasi missing data tersebut. Selanjutnya akan diteliti apakah nilai hasil imputasi missing data sudah optimal, dengan cara mencoba melakukan imputasi missing data dengan beberapa nilai k terdekat dari nilai k yang didapat dari algoritma memetika. Dari hasil imputasi akan dianalisa performansi dengan membandingkan nilai MSE hasil imputasi missing data yang didapat dari memasukkan nilai k yang didapat dari algoritma memetika dan yang didapat dari nilai k disekitar nilai k hasil optimasi algoritma memetika. Gambar 1 menunjukkan alur diagram dari penelitian.



Gambar 1 Diagram alir penelitian

2.1 Missing data

Missing data adalah data yang akan dilakukan observasi namun peneliti tidak melakukannya atau terlewat sehingga data bersangkutan terlewat atau terlupakan. Dalam setiap masalah statistika, riset, penelitian, dll, *missing data* adalah suatu masalah yang dapat menghalangi atau menghambat suatu proses penelitian. *Missing data* adalah observasi yang ingin peneliti lakukan namun tidak terjadi [7].

Missing data dapat dibagi menjadi beberapa type. *Missing Completely at Random (MCAR)*, yaitu jika distribusi data yang hilang pada suatu atribut tidak tergantung pada data pengamatan atau *missing data*. *Missing at Random (MAR)*, yaitu jika distribusi data yang hilang pada suatu atribut tergantung pada data pengamatan tetapi tidak tergantung pada *missing data*. *Not Missing at Random (NMAR)*, jika distribusi data yang hilang pada suatu atribut tergantung pada *missing data*.

2.2 Imputasi dengan metode KNN

KNN termasuk salah satu algoritma yang menggunakan pembelajaran berdasarkan pola (*pattern recognition*), yang dapat digunakan untuk proses klasifikasi dan proses regresi. *KNN* adalah metode pembelajaran pola terhadap sekumpulan data berdasarkan pembelajaran data set yang ada, yang kemudian dipelajari polanya (*training data*). Data pembelajaran kemudian diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran.

Penggunaan algoritma *KNN* dalam proses *imputasi missing data* sudah bukan hal yang baru lagi. Kelemahan pada metode *KNN* adalah jika salah memasukkan nilai k , maka algoritma ini akan memberikan output yang salah. k telah peneliti optimasi sebelumnya pada algoritma memetika. Setelah mengetahui berapa k yang seharusnya dipakai, akan dihitung berapa jarak antara subdata yang memiliki *missing data* dan data pada subdata *complete*. Untuk mengetahui, mana saja tetangga terdekat dari suatu data yang memiliki *missing data* tersebut. Berikutnya akan dikumpulkan sebanyak k tetangga terdekat dari masalah yang dicari dan dihitung rata-ratanya. Rata-rata dari kumpulan tetangga terdekat masalah tersebut akan mengisi *missing data* yang dicari.

2.3 Optimasi Algoritma Memetika

Algoritma memetika adalah pendekatan yang dilakukan berdasarkan populasi kromosom untuk pemecahan suatu masalah optimasi. Karena algoritma memetika hanya merupakan suatu pendekatan, maka algoritma tersebut termasuk kedalam metode heuristik yakni suatu metode yang digunakan ketika ukuran ruang pencarian solusi sulit dikontrol secara eksak dan belum ada algoritma yang dapat mencari solusi optimum secara efektif. Algoritma memetika memiliki kompleksitas yang lebih tinggi dibandingkan algoritma genetika (untuk domain tertentu) karena adanya penambahan solusi lokal. Komponen algoritma memetika antara lain :

2.3.1 Kromosom

Individu yang terdapat dalam satu populasi dan merupakan suatu solusi yang masih berbentuk *symbol*. Gen merupakan bagian dari kromosom, satu gen biasanya akan mewakili satu variable. Gen dapat direpresentasikan dalam bentuk : string bit, pohon, array bilangan real, daftar aturan, dan lain sebagainya. Dalam tugas akhir ini kromosom yang digunakan mewakili banyak tetangga.

2.3.2 Inisialisasi Individu

Proses mengubah data input yang dimasukkan untuk mendapatkan kromosom dalam individu tersebut, yang nantinya akan diolah untuk mendapatkan generasi berikutnya yang lebih baik dan mendapatkan solusi yang lebih baik pula. Dalam prosedur inisialisasi ini akan ditentukan ukuran populasi. Ukuran populasi tergantung pada masalah yang akan dipecahkan dan jenis operator genetika yang akan diimplementasikan. Berikutnya dilakukan inisialisasi terhadap kromosom pada populasi yang dibahas. Inisialisasi kromosom dilakukan secara acak, namun demikian harus tetap memperhatikan domain solusi dan kendala permasalahan, dalam hal ini *missing data*.

2.3.3 Nilai Fitness

Nilai fitness yaitu alat ukur yang digunakan untuk proses evaluasi kromosom. Nilai fitness dari suatu kromosom akan menunjukkan kualitas kromosom dalam populasi tersebut. Makna dari nilai fitness menandakan penalty atas pelanggaran suatu kendala. Dapat disimpulkan bahwa semakin besar nilai fitness pada suatu kromosom menandakan bahwa kromosom tersebut semakin layak menjadi solusi permasalahan. Dalam tugas akhir ini nilai fitness yang digunakan adalah MSE dari suatu individu yang diuji.

2.3.4 Crossover

Crossover (pindah silang) adalah metode yang digunakan untuk mendapatkan individu yang lebih baik. Dimisalkan 2 individu dengan kromosom biner sebagai berikut : In1 : 110 10101, In2 : 101 11101. Crossover akan menukar bagian kromosom individu 1 dengan individu 2 untuk mendapatkan individu selanjutnya yang lebih baik. Hasil crossover individu 1 dengan individu 2 akan menjadi : anak1 : 110 11101, dan anak2 : 101 10101. Apabila dalam kromosom individu tersebut bilangan biner 1 menyatakan nilai individu tersebut bagus, maka dari hasil crossover berhasil mendapatkan individu yang lebih baik yaitu anak1 : 110 11101.

2.3.5 Mutasi Gen

Mutasi gen merupakan proses pertukaran/perubahan informasi yang dibawa oleh masing-masing gen dalam kromosom anak. Mutasi ini dilakukan agar gen-gen tidak 'rusak' setelah mengalami crossover. Mutasi bertujuan untuk mengurangi kesamaan informasi yang dibawa oleh kromosom anak dengan orang tua, sehingga dapat berkontribusi menuju kromosom yang paling berkualitas. Mutasi dilakukan pada beberapa gen dalam kromosom anak dengan probabilitas mutasi kecil. Bila mengambil contoh sebelumnya anak2 tidak memiliki gen kromosom yang lebih baik dari orangtuanya, maka dari itu butuh dilakukan mutasi gen agar gen hasil crossover dapat menjadi lebih baik dari orangtuanya.

2.3.6 Pencarian lokal

Pencarian lokal adalah proses pertukaran/penempatan kembali informasi yang dibawa oleh gen-gen dalam suatu kromosom anak dengan harapan dapat meningkatkan kualitas dari kromosom tersebut. Pencarian lokal dapat dilakukan dengan menukar dua gen, atau melakukan proses mutasi pada beberapa gen tanpa mengurangi kualitas dari kromosom sebelumnya.

2.4 Metode Hybrid KNN-Memetika

Imputasi missing data dengan menggunakan KNN dengan optimasi algoritma memetika ini dilakukan dengan membagi dataset menjadi dua subdata yaitu subdata dengan instance yang memiliki atribut yang lengkap dan subdata yang mengandung minimal satu atribut yang hilang. Jika D merupakan suatu dataset yang akan diuji, maka subdata D_c adalah subdata yang komplit sedangkan D_m adalah subdata yang memiliki atribut yang hilang.

Langkah pertama dalam proses imputasi missing data ini adalah mengukur berapa calon solusi terbanyak yang mungkin diuji, dalam hal ini adalah berapa nilai k maksimal yang dapat dimasukkan. Langkah selanjutnya adalah membangkitkan kromosom k secara acak, dalam batas kemampuan penelitian. Kromosom dikodekan dengan string biner. Misalkan representasi kromosom untuk nilai $k = 7$ dapat dilihat pada Gambar .

0	0	1	1	1
---	---	---	---	---

Gambar 2 representasi kromosom

Individu yang memiliki nilai fitness paling bagus akan terpilih sebagai hasil dari optimasi algoritma memetika. Dalam hal ini individu yang didapat adalah nilai k dan nilai fitness yang dipakai diukur berdasarkan nilai MSE dari individu yang diuji. Kemudian dilakukan pengecekan dengan memasukkan nilai k yang didapat dari optimasi algoritma memetika ke dalam algoritma KNN untuk mengukur nilai MSE. Nilai imputasi missing data yang didapat dengan algoritma KNN dalam penelitian ini adalah dengan cara menjumlahkan sebanyak k data tetangga diatas dan dibawah nilai missing data yang kemudian dibagi dengan jumlah element data yang digunakan. Digambarkan dengan persamaan berikut :

$$\frac{\sum}{n} \quad (1)$$

Dengan x mewakili nilai imputasi missing data dan n mewakili jumlah data yang digunakan. Dan nilai MSE dapat ditentukan dengan persamaan berikut :

$$\frac{\sum}{n} \quad (2)$$

3 Pembahasan

3.1 Dataset

Dari data set lengkap yang didapat, akan dihilangkan data secara acak. Kemudian data dapat dibagi menjadi 2 subdata, data yang berisi *missing data* didalamnya dan data yang masih lengkap. Data-data tersebut kemudian akan dilakukan inialisasi, atau pembangkitan kromosom pada masing masing individu. Terdapat beberapa macam jenis kromosom, seperti kromosom biner, kromosom float, kromosom string, dan kromosom kombinatorial. Dalam penelitian tugas akhir ini tipe kromosom yang akan dipakai adalah tipe kromosom bilangan biner. Karena kromosom biner merupakan model yang sederhana dengan tingkat keberhasilan yang tinggi. Bentuk kromosom biner ini juga disarankan untuk digunakan dalam permasalahan dengan parameter yang jumlah dan rangenya tertentu.

3.2 Hasil Pengujian dan Pembahasan

Pada scenario ini, akan dibuat missing data 5%, 10%, 15%, 20% dan 25% pada dataset Abalone. Kemudian akan dilakukan prosedur seperti pada rancangan aplikasi, memasukkan k hasil optimasi memetika untuk imputasi KNN. Selanjutnya akan dicari juga beberapa nilai k lebih rendah dan lebih tinggi, sehingga dapat diketahui apakah nilai k yang didapat dari optimasi algoritma memetika sudah optimal.

Missing data sebanyak 5%:

Tabel 1 Tabel Parameter Pengujian 1

K	MSE	Keterangan
1	6.8535	KNN
2	6.6203	KNN
3	6.8432	KNN dioptimasi Memetika
4	6.8782	KNN
5	7.2132	KNN
6	7.234	KNN
7	7.709	KNN
8	7.7664	KNN
9	8.2027	KNN

Hasil pengujian pada skenario pembangkitan missing data sebanyak 5%. Optimasi algoritma memetika menemukan, nilai k optimal pada saat k memiliki nilai 3 dengan nilai MSE sebesar 6.8432. Namun setelah diteliti ditemukan nilai MSE lebih kecil dengan nilai 6.6203 pada saat k memiliki nilai 2.

Missing data sebanyak 10% :

Tabel 2 Tabel Parameter Pengujian 2

K	MSE	Keterangan
1	17.3184	KNN
2	16.5493	KNN
3	16.3693	KNN
4	16.5621	KNN dioptimasi Memetika
5	17.2051	KNN
6	17.4913	KNN
7	17.9396	KNN
8	18.2594	KNN
9	19.2215	KNN

Hasil pengujian pada skenario pembangkitan missing data sebanyak 10%. Optimasi algoritma memetika menemukan, nilai k optimal pada saat k memiliki nilai 4 dengan nilai MSE sebesar 16.5621. Namun setelah diteliti ditemukan nilai MSE lebih kecil dengan nilai 16.3693 pada saat k memiliki nilai 3.

Missing data sebanyak 15%

Tabel 3 Tabel Parameter Pengujian 3

K	MSE	Keterangan
---	-----	------------

1	24.5512	KNN
2	23.6767	KNN dioptimasi Memetika
3	23.9215	KNN
4	24.4364	KNN
5	25.993	KNN
6	26.2802	KNN
7	28.0225	KNN
8	28.8808	KNN
9	29.9752	KNN

Hasil pengujian pada skenario pembangkitan missing data sebanyak 15%. Optimasi algoritma memetika menemukan, nilai k optimal pada saat k memiliki nilai 2 dengan nilai MSE sebesar 23.6767. Dengan tidak ada nilai MSE lain yang lebih kecil.

Missing data sebanyak 20%

Tabel 4 Tabel Parameter Pengujian 4

K	MSE	Keterangan
1	30.1838	KNN
2	29.9327	KNN
3	29.828	KNN
4	29.9201	KNN dioptimasi Memetika
5	31.1907	KNN
6	32.5459	KNN
7	33.8049	KNN
8	35.1405	KNN
9	36.7179	KNN

Hasil pengujian pada skenario pembangkitan missing data sebanyak 20%. Optimasi algoritma memetika menemukan, nilai k optimal pada saat k memiliki nilai 4 dengan nilai MSE sebesar 29.9201. Namun setelah diteliti ditemukan nilai MSE lebih kecil dengan nilai 29.828 pada saat k memiliki nilai 3.

Missing data sebanyak 25%

Tabel 5 Tabel Parameter Pengujian 5

K	MSE	Keterangan
1	40.934	KNN
2	37.8428	KNN dioptimasi Memetika
3	37.5737	KNN
4	38.3904	KNN
5	40.2173	KNN
6	40.85	KNN
7	42.7324	KNN

8	44.1677	KNN
9	45.8731	KNN

Hasil pengujian pada skenario pembangkitan missing data sebanyak 25%. Optimasi algoritma memetika menemukan, nilai k optimal pada saat k memiliki nilai 2 dengan nilai MSE sebesar 37.8428. Namun setelah diteliti ditemukan nilai MSE lebih kecil dengan nilai 37.5737 pada saat k memiliki nilai 3.

Dari hasil pengujian tertulis tersebut dapat dianalisis sebagai berikut, nilai k yang didapat dari hasil optimasi algoritma memetika tidak selalu merupakan nilai MSE terkecil yang ada. Berdasarkan data yang diperoleh, nilai MSE yang didapatkan dalam dari memasukkan nilai k optimal tidak terlalu jauh berbeda dengan nilai MSE optimal yang ada. Dan meningkatnya nilai error seiring dengan makin banyaknya porsi data yang hilang.

Tabel 6 Tabel Rangkuman Hasil Pengujian

%	Optimal Pada	k	MSE
5	KNN dioptimasi Memetika	3	6.8432
	KNN	2	6.6203
10	KNN dioptimasi Memetika	4	16.5621
	KNN	3	16.3693
15	KNN dioptimasi Memetika	2	23.6767
	KNN	2	23.6767
20	KNN dioptimasi Memetika	4	29.9201
	KNN	3	29.828
25	KNN dioptimasi Memetika	2	37.8428
	KNN	3	37.5737

4 Kesimpulan dan Saran

Aplikasi imputasi missing data dengan metode KNN yang dioptimasi oleh algoritma memetika dapat melakukan imputasi missing data dengan tingkat kesalahan yang relatif kecil. Bergantung dari alat pengukur akurasi dalam imputasi missing data. Nilai k yang diperoleh dengan algoritma memetika tidak selalu menghasilkan nilai k yang optimal. Namun nilai k yang diperoleh sangat mendekati hasil optimal yang diharapkan. Banyaknya nilai missing data yang ada mempengaruhi tingkat akurasi imputasi missing data juga ikut berubah. Semakin banyak missing data maka akan semakin menurun tingkat akurasi. Melakukan beberapa perubahan pada dataset tanpa mengubah nilai dari dataset dapat mempengaruhi nilai akurasi imputasi missing data.

Saran yang dapat diberikan untuk pengembangan *sistem* ini antara lain. Dapat memeriksa akurasi k disekitar k optimal yang diperoleh dari hasil optimasi algoritma memetika atau algoritma optimasi lainnya, untuk memberikan informasi k yang diperoleh dari hasil optimasi adalah k optimal dari suatu cara pengukuran akurasi tersebut. Jika tidak, maka k yang memiliki akurasi lebih bagus akan dipakai sebagai k optimal. Dapat dilakukan percobaan dengan data yang lebih beragam dan algoritma optimasi lainnya, agar diperoleh algoritma alternatif dalam menentukan k optimal dalam KNN. Dapat ditambah metode pengukuran akurasi lain untuk memberikan gambaran lebih jelas seberapa dekat hasil *imputasi* dengan nilai yang diharapkan.

Daftar Pustaka

- [1] Izzah, Abidatul., 2012. *Imputasi Missing data menggunakan Metode K-Nearest Neighbour Dengan Optimasi Algoritma Genetika*. Jurnal. Surabaya: ITS.
- [2] XU, Jin., dkk., 2014. *A Memetic Algorithm Based on k-Nearest Neighbour*. Journal of Computational Information Systems. China: Hubei University.
- [3] Garg, Poonam., 2009. *A Comparison Between Memetic Algorithm and Genetic Algorithm for the Cryptanalysis of Simplified Data Encryption Standard Algorithm*. International Journal of Network Security & Its Applications. India: Institute of Management Technology.
- [4] Rubin, Donald B., 1976. *Inference and Missind Data*. Biometrika, Vol. 63, No.3, 581-592.
- [5] Kumar, Rakesh., Tyagi, Sanjay., Sharma, Manju., 2013. *Memetic Algorithm: Hybridization of Hill Climbing with Selection Operator*. International Journal of Soft Computing and Engineering. India: Kurukshetra University.
- [6] Haryono, Riendu Magistriyanti., 2012. *Penjadwalan Ujian Menggunakan Algoritma Memetika Dengan Pengkodean Langsung*. Tugas Akhir. Bandung: IT Telkom.
- [7] <http://missingdata.lshtm.ac.uk>. Carpenter, James., dkk 2004 (Diakses pada 6 Februari 2015 Pk 15.20)
- [8] http://www.saedsayad.com/k_nearest_neighbors_reg.htm. Sayad, Saed. 2010. *K Nearest Neighbour - Regresion*. (Diakses pada 6 Februari 2015 Pk 20.43)
- [9] Parvin, Hamid., dkk., 2010. *A Modification on K-Nearest Neighbor Classifier*. Global Journal of Computer Science and Technology. Iran: GJCST Classification.
- [10] Suguna, N., Thanushkodi, T., 2010. *An Improved k-Nearest Neighbor Classification Using Genetic Algorithm*. International Journal of Computer Science Issues. India: Akshaya College of Engineering and Technology.