

SISTEM DIAGNOSA KESUBURAN SPERMA DENGAN METODE K-NEAREST NEIGHBOR (K-NN)

Rizal Yepriyanto¹⁾, Kustanto²⁾, Yustina Retno Wahyu Utami³⁾

Abstract

Patterns and lifestyle of people today lead a lot discovered problems or diseases suffered by humans that can damage the health. One of these sperm fertility problems in men who have declined in the last two decades. It was explained that environmental factors and lifestyle may affect sperm quality. Several factors can affect sperm fertility include age, high fever, smoking habits and alcohol consumption. It needs to make an application that can help diagnose early fertility of sperm. The process starts from the search criteria and make the training data by taking the results of the UCI Repository. The next step is the calculation of the K-NN method that gives results in the form of sperm diagnosis. Results of this research is a website application interface, and the results of testing the accuracy of the data K-NN performed with 60 training data and testing the data 20, obtained accuracy value by 85%.

Keywords: Fertility Sperm, Diagnosis, K-NN, Website

I. PENDAHULUAN

Angka infertilitas pasangan suami-istri di Indonesia yang mengalami kesulitan untuk mendapatkan anak sekitar 10%. Faktor dari pihak pria masih merupakan penyebab terpenting dari infertilitas. WHO mendapatkan bahwa lebih dari 50% penyebab infertilitas adalah pihak pria dan yang terbesar oleh karena faktor semen. Ada beberapa faktor yang dapat mempengaruhi kesuburan sperma salah satunya yaitu pola gaya hidup dan lingkungan (David dkk, 2012). Pola gaya hidup dan lingkungan tertentu memiliki pengaruh terhadap konsentrasi sperma, motilitas sperma (gerak aktif) dan morfologi sperma (bentuk normal) yang merupakan faktor yang dapat mempengaruhi kualitas sperma. Kriteria atau parameter yang digunakan dalam menentukan kesuburan sperma berdasarkan data

^{1,2,3)} Program Studi Teknik Informatika, STMIK Sinar Nusantara Surakarta

sampel yang diambil dari *UCI Machine Learning Repository* yang berjudul “*Fertility Data Set*” (David dkk, 2012). Jurnal penelitian yang berjudul “*Predicting Seminal Quality with Artificial Intelligence Methods*” (David dkk, 2012) merupakan penelitian yang dilakukan dengan data sampel yang sama. Dalam proses prediksi kualitas sperma, penelitian menggunakan faktor pola gaya hidup dan lingkungan sebagai acuan dalam menentukan kualitas sperma. Hasil klasifikasi atau output yang diberikan yaitu *Normal* dan *Altered*.

Metode algoritma yang digunakan dalam penelitian menggunakan algoritma klasifikasi yaitu *K-Nearest Neighbor* (K-NN), dengan menggunakan data sampel yang sama dengan penelitian terdahulu. Penggunaan algoritma *K-Nearest Neighbor* (K-NN) didasarkan pada penelitian berkaitan dengan klasifikasi data dari beberapa kriteria dan jumlah data sampel yang digunakan cukup banyak, karena metode K-NN akan menghasilkan akurasi yang baik pada data yang berjumlah banyak dan memiliki banyak label kelas.

Berdasarkan permasalahan tersebut, dibuatlah Sistem Diagnosa Kesuburan Sperma Dengan Metode *K-Nearest Neighbor* (K-NN) yang diharapkan dapat digunakan untuk deteksi dini kesuburan sperma.

II. TINJAUAN PUSTAKA

1.1 Kecerdasan Buatan

Kecerdasan Buatan (*Artificial Intelligence*) adalah ilmu pengetahuan yang mempelajari dan menerapkan kemampuan berpikir manusia dan pengetahuan sebagai pengolah proses informasi dan metode penelitian ilmu pengetahuan. Kecerdasan buatan merupakan studi tentang bagaimana membuat komputer melakukan hal-hal yang pada saat ini dapat dilakukan lebih baik oleh manusia (Struart, 2010).

1.2 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar. Menurut Gartner Group, data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan

dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2006).

1.3 K-Nearest Neighbour (K-NN)

Prinsip kerja *k-Nearest Neighbour* (*k-NN*) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan *k* tetangga (*Neighbour*) terdekatnya dalam data pelatihan. Berikut urutan proses kerja *k-NN* (Gorunescu, 2011):

1. Menentukan parameter *k* (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak *euclidean* (*euclidean distance*) masing-masing obyek terhadap data sampel yang diberikan.

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

Keterangan:

x_1 = Sampel Data

x_2 = Data Uji / Testing

i = Variabel Data

d = Jarak

p = Dimensi Data

3. Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclidean terkecil
4. Mengumpulkan kategori *y* (klasifikasi *nearest neighbor*)

1.4 Spermatozoid

Spermatozoid atau sel sperma atau *spermatozoa* (berasal dari bahasa Yunani kuno: σπέρμα yang berarti benih, dan ζῷον yang berarti makhluk hidup) adalah sel dari sistem reproduksi laki-laki. Sel sperma akan membuahi *ovum* untuk membentuk *zigot*. *Zigot* adalah sebuah sel dengan kromosom lengkap yang akan berkembang menjadi embrio.

Saat mencapai pubertas, testis pria mulai mengalami proses *spermatogenesis*. Pada pria, sel-sel yang menjalani meiosis disebut “*spermatosit*”. *Spermatosit* primer yang menjalani pembelahan meiosis pertama berisi kromosom-kromosom dalam jumlah diploid (dua pasang). Tiap *spermatosit* sekunder mempunyai 22 otosom dan satu kromosom seks. Satu kromosom berisi kromosom X (ditambah salinannya) dan satu kromosom Y (ditambah salinannya) (Jannah, 2011).

1.5 Diagram Alir Data (DAD)

Diagram Alir Data (DAD) atau *Data Flow Diagram* (DFD) adalah representasi grafik yang menggambarkan aliran informasi dan transformasi informasi yang diaplikasikan sebagai data yang mengalir dari masukan (*input*) dan keluaran (*output*) (Rosa dan Shalahuddin, 2011).

1.6 PHP (*Hypertext Preprocessor*)

PHP adalah bahasa skrip yang dapat ditanamkan atau disisipkan ke dalam HTML. PHP banyak dipakai untuk memprogram situs web dinamis. Pada awalnya PHP merupakan kependekan dari *Personal Home Page* (Situs personal). PHP pertama kali dibuat oleh Rasmus Lerdorf pada tahun 1995. Pada waktu itu PHP masih bernama *Form Interpreted* (FI), yang wujudnya berupa sekumpulan skrip yang digunakan untuk mengolah data formulir dari web (Aditya, 2011).

1.7 MySQL

MySQL adalah sebuah perangkat lunak sistem manajemen basis data SQL atau DBMS (*Database Management System*) yang *multi-thread* dan *multi-user* dengan sekitar 6 juta instalasi di seluruh dunia. MySQL tersedia sebagai perangkat lunak gratis dibawah lisensi GNU *General Public License* (GPL), tetapi mereka juga menjual di bawah lisensi komersial untuk kasus-kasus dimana penggunaannya tidak cocok dengan penggunaan GPL (Aditya, 2011).

III. METODE PENELITIAN

a. Metode Pengumpulan Data

2. Studi Pustaka

Studi pustaka yang akan dilakukan yaitu melakukan studi metode *K-Nearest Neighbor*. Materi yang akan digunakan dan mengenai kesuburan sperma melalui literatur-literatur seperti buku, jurnal ilmiah, thesis dan sumber lain yang berkaitan dengan penelitian ini.

3. Browsing Internet

Pengumpulan data dengan menggunakan media internet dalam mencari artikel dari berita online yang memuat berita mengenai hal apa saja yang mempengaruhi kesuburan sperma. Artikel berita berasal dari beberapa media berita online seperti malezones.com, health.detik.com, bidanku.com, dan

health.liputan6.com. Data sampel digunakan yang berasal dari Website UCI Machine Learning Repository (<http://archive.ics.uci.edu>).

4. Sumber Data

Sumber data yang diperlukan untuk penelitian ini didapatkan dari data set yang berasal dari UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/index.html>) yang diterbitkan pada tahun 2013.

b. Metode Pengembangan Sistem

1. Analisis Sistem yang Diusulkan

Pada tahap ini dilakukan analisis sistem untuk diagnosa kesuburan sperma dengan metode K-Nearest Neighbor (K-NN).

2. Perancangan Sistem

Perancangan sistem menggunakan HIPO, Diagram Konteks dan DAD (Diagram Alir Data).

3. Implementasi

Implementasi sistem berbasis web menggunakan pemrograman PHP dan database menggunakan MySQL.

4. Pengujian

Pengujian fungsional sistem menggunakan metode *Black Box* dan Pengujian validitas didasarkan pada perbandingan antara hasil data sampel dan data hasil dari perhitungan sistem menggunakan metode *K-NN*.

IV. HASIL DAN PEMBAHASAN

4.1 Analisis Data Sampel

Data sampel atau data set yang digunakan bersumber dari UCI Machine Learning Repository. Data sampel yang digunakan berjumlah 60 record. Atribut/kriteria yang digunakan yaitu:

- Musim saat analisis dilakukan
- Usia pada saat analisis
- Penyakit masa kekanakan (misalnya, cacar air, campak, gondok, polio)
- Kecelakaan atau trauma serius
- Tindakan bedah
- Demam tinggi dalam setahun terakhir
- Frekuensi konsumsi alkohol

- Kebiasaan merokok
- Lamanya jam yang dihabiskan untuk duduk per hari

Dari beberapa parameter/variabel diatas dapat ditentukan hasil diagnosa kesuburan sperma. Sistem diagnosa kesuburan sperma memiliki 2 hasil analisis yaitu:

1. Normal

Hasil diagnosa “*Normal*” berarti sperma memiliki karakteristik normal atau “*Normozoospermia*”. Jadi jumlah sel *spermatozoa*, gerakan, dan bentuknya normal atau dengan kata lain kesuburan sperma normal.

2. Altered

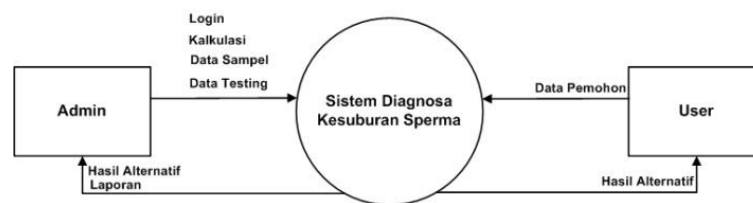
Hasil diagnosa “*Altered*” berarti sperma memiliki perubahan karakteristik. Hasil diagnosa “*Altered*” menunjukkan bahwa karakteristik sperma tidak normal dengan kata lain sperma mengalami ketidaksuburan.

Data training yang digunakan dalam perhitungan *k-Nearest Neighbour* dapat dilihat pada Tabel 5.

4.2 Perancangan Sistem

1. *Context Diagram*

Context Diagram pada Sistem Diagnosa Kesuburan Sperma ini menggambarkan sistem aliran program secara umum, adapun gambaran *context diagram* adalah terlihat pada Gambar 1.



Gambar 1. *Context Diagram*

2. *Data Flow Diagram (DFD) level 0*

DFD level 0 merupakan penjabaran dari *context diagram* yaitu pada prosesnya lebih diperinci, seperti terlihat pada Gambar 2.

Tabel 5. Data Training

A	B	C	D	E	F	G	H	I	J		A	B	C	D	E	F	G	H	I	J
-0,33	0,78	1	0	0	1	1	1	0,06	O		-0,33	0,92	1	1	0	1	1	-1	0,63	N
1	0,69	1	0	1	0	0,8	-1	0,25	O		-0,33	0,53	1	1	0	0	0,8	0	0,75	N
-0,33	0,94	1	0	1	0	0,8	1	0,31	O		-0,33	0,69	1	1	1	-1	1	-1	0,75	N
1	0,67	1	0	0	0	0,8	1	0,38	O		-0,33	0,5	1	1	0	-1	0,8	0	0,88	N
1	0,67	1	0	1	0	0,6	-1	0,38	O		-0,33	0,69	0	1	1	0	0,8	0	0,88	N
1	0,78	1	1	0	1	0,6	-1	0,38	O		1	0,56	0	0	0	1	0,8	0	1	N
1	0,69	1	0	1	-1	1	-1	0,44	O		-1	0,53	1	1	0	1	1	0	0,31	N
0,33	0,75	1	0	1	0	0,8	-1	0,44	O		-1	0,72	1	1	1	1	0,8	-1	0,19	N
-0,33	0,67	1	1	0	0	0,8	-1	0,5	O		1	0,75	1	1	1	0	1	1	0,25	N
1	0,67	0	0	1	0	0,6	0	0,5	O		-0,33	0,81	1	1	1	0	1	1	0,38	N
-1	0,67	0	0	1	0	0,6	0	0,5	O		-0,33	0,83	1	1	1	0	1	-1	0,31	N
-0,33	0,5	1	1	0	-1	0,8	0	0,88	O		-0,33	0,89	1	1	0	0	0,6	1	0,31	N
0,33	0,78	1	0	0	0	1	1	0,06	N		0,33	0,78	1	0	1	1	1	1	0,13	C
1	0,78	1	1	1	0	0,6	0	0,13	N		1	0,69	0	0	1	0	0,8	1	0,31	C
1	0,58	0	0	1	0	1	-1	0,19	N		-0,33	0,97	1	0	1	0	0,8	1	0,25	C
-0,33	0,58	1	1	1	-1	0,8	0	0,19	N		1	0,67	1	0	1	0	0,8	1	0,38	C
-0,33	0,58	1	0	1	0	0,8	1	0,19	N		0,33	0,67	1	0	1	0	0,8	-1	0,38	C
-1	0,64	1	0	1	0	1	0	0,19	N		1	0,78	1	1	0	1	0,6	-1	0,44	C
-0,33	0,67	1	0	1	0	0,8	-1	0,19	N		1	0,69	1	1	1	-1	1	-1	0,38	C
-1	0,78	1	0	1	1	1	1	0,25	N		1	0,78	1	1	0	0	0,8	-1	0,44	C
-0,33	0,86	1	1	1	1	1	-1	0,25	N		-0,33	0,67	1	1	0	0	0,6	-1	0,5	C
1	0,94	1	1	1	0	0,2	-1	0,25	N		1	0,67	0	0	1	0	0,6	0	0,56	C
-1	1	1	0	1	1	0,6	0	0,25	N		-1	0,67	0	0	1	0	0,8	0	0,5	C
-0,33	0,5	1	0	0	1	0,8	0	0,31	N		0,33	0,5	0	1	1	-1	0,8	0	0,81	C
-1	0,56	1	0	1	0	1	-1	0,63	N		0,33	0,78	1	1	0	0	1	1	0,06	C
-0,33	0,56	1	0	0	0	1	-1	0,63	N		1	0,69	1	0	1	1	0,8	-1	0,25	C
-0,33	0,56	1	1	0	0	0,4	1	0,63	N		-0,33	0,94	1	1	1	0	0,8	1	0,31	C
1	0,61	1	0	1	0	1	-1	0,63	N		1	0,67	1	0	0	0	0,6	1	0,38	C
-0,33	0,61	1	0	1	0	1	-1	0,63	N		1	0,67	1	0	1	1	0,6	-1	0,38	C
1	0,64	0	0	0	0	1	-1	0,63	N		-0,33	0,67	1	1	0	0	0,8	-1	0,44	C

Keterangan :

A = Musim saat analisis

B = Umur (Tahun)

C = Penyakit masa kanak

D = Kecelakaan/trauma

E = Tindakan bedah

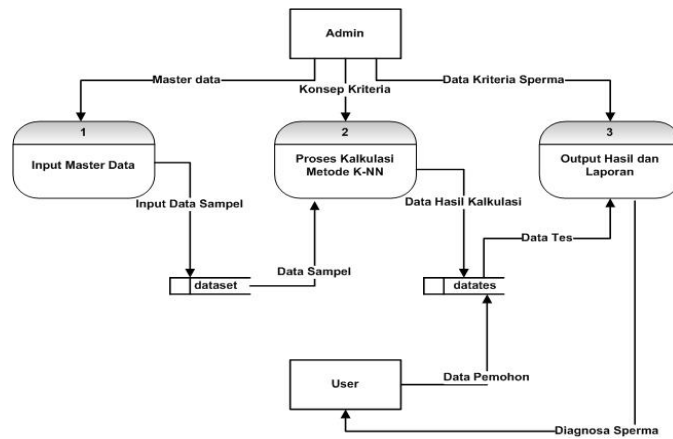
F = Demam tinggi

G = Kebiasaan merokok

H = Frekuensi konsumsi alkohol

I = Lama duduk yang dihabiskan per hari (Jam)

J = Hasil diagnosa (Output)



Gambar 2. DFD level 0

4.3 Perhitungan Data

Data testing diambil dari UCI Machine Learning Repository. Data testing/uji yang digunakan dalam perhitungan k-NN berjumlah 20 record seperti pada Tabel 6.

Tabel 6. Data Testing

T	A	B	C	D	E	F	G	H	I	J
T1	-0,33	0,75	1	1	1	0	0,6	-1	0,19	?
T2	-1	0,81	1	1	1	1	0,8	0	0,19	?
T3	-1	0,64	1	0	0	1	1	1	0,25	?
T4	-1	0,58	1	0	1	-1	0,8	1	0,5	?
T5	-1	0,53	1	0	0	1	0,8	-1	0,31	?
T6	-0,33	0,64	1	1	1	0	0,8	-1	0,31	?
T7	1	0,67	1	1	0	-1	0,8	0	0,31	?
T8	1	0,64	1	0	1	0	1	-1	0,38	?
T9	-0,33	0,75	0	1	1	0	1	-1	0,38	?
T10	-0,33	0,92	1	0	0	1	0,6	-1	0,19	?
T11	-0,33	0,67	0	0	0	-1	0,8	-1	0,44	?
T12	-0,33	0,5	1	0	0	1	1	-1	0,47	?
T13	-1	0,75	1	0	1	1	0,6	0	0,56	?
T14	1	0,67	0	1	1	0	0,6	-1	0,38	?
T15	0,33	0,5	0	1	0	-1	0,8	0	0,88	?
T16	1	0,69	1	0	1	0	1	-1	0,38	?
T17	1	0,75	1	1	1	0	0,6	-1	0,25	?
T18	-0,33	0,78	0	1	1	1	1	1	0,19	?
T19	1	0,69	1	0	1	0	1	1	0,25	?
T20	-0,33	0,94	1	0	0	1	0,8	1	0,31	?

Keterangan :

A= Musim saat analisis	F = Demam tinggi	I = Lama duduk yang
B = Umur (Tahun)	G = Kebiasaan merokok	dihabiskan per hari (Jam)
C = Penyakit masa kanak	H = Frekuensi konsumsi	J = Hasil diagnosa (Output)
D = Kecelakaan/trauma	alkohol	T = Data testing
E = Tindakan bedah		

Tahap perhitungan *K-NN* yang pertama adalah menentukan jarak *euclidian* dari data yang dituju terhadap data training yang sudah ada. Contoh perhitungan data testing T1:

$$d_{T1} = \sqrt{\begin{matrix} ((-0,33) - (-0,33))^2 + (0,78 - 0,75)^2 + (1 - 1)^2 + \\ (0 - 1)^2 + (0 - 1)^2 + (1 - 0)^2 + (1 - 0,6)^2 + \\ (1 - (-1))^2 + (0,06 - 0,19)^2 \end{matrix}} \\ = 2,679141654$$

Setelah mengetahui jarak euclidian dari data testing T1 terhadap data training maka dapat dilakukan perangkingan dari jarak terdekat ke jarak terjauh. Setelah dilakukan perhitungan jarak dengan nilai $k = 7$, maka akan diambil tujuh record data training yang memiliki jarak terdekat. Mayoritas memiliki “output = N” yaitu data training yang memiliki “output = N” berjumlah 4 record data sedangkan data training yang memiliki “output = O” hanya berjumlah 3 record data. Hasil akhir menunjukkan bahwa data testing T1 memiliki “output = N (Normal)”.

4.4 Implementasi

Implementasi progam berisi paparan gambar asli dari program sistem diagnosa kesuburan sperma. Langkah pertama dalam membuka halaman beranda adalah membuka web browser, kemudian pada address bar ketikan <http://localhost/sperma>, setelah ditekan enter atau klik icon panah pada browser maka akan tampil halaman beranda, kemudian pilih menu pengujian data dan klik pengujian data baru maka akan tampil halaman untuk memasukkan beberapa kriteria untuk selanjutnya perhitungan akan diproses oleh sistem agar menghasilkan hasil diagnosa, terlihat pada Gambar 3.

Setelah mengisi data kriteria klik tombol simpan maka akan keluar sebuah halaman hasil perhitungan yang menampilkan hasil diagnosa sperma seperti pada Gambar 4.

Gambar 3. Halaman Input Data Testing

Gambar 4. Halaman Hasil Pengujian

4.5 Pengujian

Sistem dikatakan memiliki akurasi yang tinggi apabila output/hasil klasifikasi yang ditentukan dengan metode k -NN memiliki nilai yang sama dengan status atau hasil klasifikasi pada data sampel. Berikut merupakan hasil keseluruhan pengujian validitas sistem dengan program k -NN dan hasil data sampel

Tabel 7. Perbandingan Hasil Klasifikasi

D	Hasil Pengujian		Keterangan	D	Hasil Pengujian		Keterangan
	Program k -NN	Data Sampel			Program k -NN	Data Sampel	
D1	N	N	T	D11	N	N	T
D2	N	N	T	D12	N	N	T
D3	N	N	T	D13	N	N	T
D4	N	N	T	D14	O	O	T
D5	N	N	T	D15	N	O	F
D6	N	N	T	D16	O	O	T

D	Hasil Pengujian		Keterangan	D	Hasil Pengujian		Keterangan
	Program k-NN	Data Sampel			Program k-NN	Data Sampel	
D7	O	N	F	D17	O	O	T
D8	O	N	F	D18	O	O	T
D9	N	N	T	D19	O	O	T
D10	N	N	T	D20	O	O	T

Keterangan:

D = Data Sampel

T = True. Terjadi apabila hasil data sampel sama dengan hasil program.

F = False. Terjadi apabila hasil data sampel berbeda dengan hasil program

Berdasarkan pengujian validitas yang telah dilakukan maka diperoleh:

$$\begin{aligned}
 \text{Akurasi} &= \frac{\sum \text{data uji benar}}{\sum \text{jumlah total data uji}} \times 100\% \\
 &= \frac{17}{20} \times 100\% = 85\%
 \end{aligned}$$

Hasil pengujian menunjukkan bahwa kinerja sistem sudah cukup baik dan layak karena akurasi mencapai 85%. Dimungkinkan akurasi dapat meningkat jika data sampel yang digunakan lebih banyak. Data sampel dapat ditambah seiring dengan seringnya program aplikasi ini digunakan.

V. KESIMPULAN

Berdasarkan penerapan metode *k-Nearest Neighbour* pada Sistem Diagnosa Kesuburan Sperma dapat diambil kesimpulan sebagai berikut:

1. Terciptanya sebuah sistem diagnosa kesuburan sperma dengan menggunakan metode *k-Nearest Neighbour* (*k*-NN).
2. Berdasarkan pengujian dengan 20 data uji, 17 data terklasifikasi dengan benar, maka kinerja atau akurasi sistem adalah 85%. Dimungkinkan akurasi dapat meningkat jika data sampel yang digunakan lebih banyak.

DAFTAR PUSTAKA

- Aditya, Nur Alan, *Jago PHP dan MySQL*, Dunia Komputer, Bekasi, 2011.
- David, G., Jose Luis Girela, Joaquin De Juan, M. Jose Gomez-Torrez, Magnus Johnsson, 2012, “*Predicting Seminal Quality with Artificial Intelligence Methods*”. Expert Systems with Applications, http://www.researchgate.net/publication/230868076_Predicting_seminal_quality_with_artificial_intelligence_methods/file/79e415058f10cc3081.pdf, (Diakses 17 Maret 2014).
- David, G, Jose Luis Girela. 2013. *Fertility Data Set*, <http://archive.ics.uci.edu/ml/datasets/Fertility#>, (Diakses 16 Maret 2014).
- Gunawan Kuswondo, *Analisis Semen Pada Pasangan Infertil*, Program Pendidikan Dokter Spesialis I Obstetri Dan Ginekologi, Universitas Diponegoro, Semarang, 2002.
- Gorunescu, Florin, *Data Mining: Concepts, Models and Techniques*, Springer, 2011.
- Jannah, Nurul, *Biologi Reproduksi*, Ar-Ruzz Media, Yogyakarta, 2011.
- Larose, Daniel T., *Data Mining Methods And Models*, John Wiley & Sons, Inc, New Jersey, 2006.
- Rosa AS, M. Shalahuddin, *Modul Pembelajaran Rekayasa Perangkat Lunak*, Modula, Bandung, 2011.
- Stuart, Russell, Peter Norvig, *Artifial Intelligence a Modern Approach Third Edition*, Pearson Education, New Jersey, 2010.