

**PENERAPAN *FUZZY K-NEAREST NEIGHBOR* UNTUK
MENENTUKAN STATUS EVALUASI KINERJA KARYAWAN**

Untuk memenuhi sebagian persyaratan mencapai gelar Sarjana Komputer



Disusun oleh :

YANITA CAHYANINGTYAS

NIM. 0810963075

**PROGRAM STUDI INFORMATIKA / ILMU KOMPUTER
PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2013**

PENERAPAN *FUZZY K-NEAREST NEIGHBOR* UNTUK MENENTUKAN STATUS EVALUASI KINERJA KARYAWAN

Yanita Cahyaningtyas¹, Achmad Ridok², Candra Dewi³

Jurusan Ilmu Komputer Program Teknologi Informasi dan Ilmu Komputer
Universitas Brawijaya

Jalan MT. Haryono 167, Malang 65145, Indonesia

1. E-mail: ninit.gaga@gmail.com
2. E-mail: acridok@ub.ac.id

Abstrak

Sumber daya manusia yang mempunyai kompetensi tinggi sangat diperlukan dalam suatu perusahaan. Dengan keahlian yang dimiliki akan menjadi pertimbangan bagi suatu perusahaan untuk melakukan evaluasi kinerja. Evaluasi kinerja karyawan dilihat dari aspek-aspek yang telah ditentukan perusahaan dimana karyawan yang dievaluasi akan mempunyai nilai evaluasi yang berbeda-beda untuk setiap aspeknya. Dengan nilai evaluasi yang berbeda-beda tersebut akan menjadi suatu masalah apabila dikaitkan dengan hasil keputusan yang akan berguna untuk mengembangkan sumber daya manusia dalam suatu perusahaan. Maka diperlukan suatu penggalan data untuk menentukan status evaluasi kinerja karyawan pada evaluasi kinerja karyawan. Dalam penelitian ini digunakan metode *Fuzzy K-Nearest Neighbor* untuk menentukan status evaluasi kinerja karyawan. Proses yang dilakukan untuk menentukan status evaluasi kinerja karyawan meliputi dua tahap, yaitu tahap normalisasi nilai atribut dan tahap klasifikasi dengan algoritma fuzzy k-nearest neighbor. Tahap normalisasi dilakukan untuk menormalisasikan nilai atribut pada dataset. Setelah didapatkan nilai atribut yang ternormalisasi dilakukan tahap klasifikasi yang digunakan untuk mencari distance yang diperlukan untuk mendapatkan nilai derajat keanggotaan dimana dipilih nilai derajat keanggotaan terbesar yang menentukan didapatkannya kelas target. Hasil penelitian menunjukkan bahwa semakin bertambahnya nilai tetangga terdekat maka akurasi yang didapatkan semakin menurun dan stabil pada nilai tetangga terdekat tertentu untuk masing-masing data latih. Sedangkan untuk nilai tetangga terdekat yang kecil dapat mendominasi dataset yang disebabkan oleh kelas yang dominan pada data tersebut.

Kata kunci : *Fuzzy K-Nearest Neighbor*, *Data Mining*, klasifikasi, evaluasi kinerja karyawan.

I. PENDAHULUAN

Sumber daya manusia yang mempunyai kompetensi tinggi sangat diperlukan karena keahlian atau kompetensi akan dapat mendukung peningkatan prestasi kerja karyawan. Dalam suatu perusahaan diperlukan sumber daya manusia yang berkualitas dimana menjadi salah satu faktor yang digunakan untuk meningkatkan kinerja[11]. Kinerja karyawan merupakan prestasi kerja atau hasil kerja yang dicapai sumber daya manusia persatuan periode waktu dalam melaksanakan tugas kerjanya sesuai dengan tanggung jawab yang diberikan kepadanya[5]. Suatu kinerja individu dapat ditingkatkan apabila ada kesesuaian antara pekerjaan dan kemampuan. Kinerja individu sendiri dipengaruhi oleh kepuasan kerja. Kepuasan kerja itu sendiri adalah perasaan

individu terhadap pekerjaannya dimana perasaan itu berupa hasil penilaian mengenai seberapa jauh pekerjaannya secara keseluruhan. Dalam hal ini dibutuhkan suatu evaluasi, yang kemudian dikenal dengan penilaian kinerja.

Penilaian kinerja harus dilakukan untuk mengetahui prestasi yang dapat dicapai setiap karyawan. Proses evaluasi kinerja penting bagi setiap karyawan yang harus dilakukan secara adil, jujur, relevan dan valid yang sesuai dengan pekerjaan yang dilakukan serta bebas dari hal diskriminasi karena hal ini berguna bagi perusahaan untuk menetapkan tindakan kebijaksanaan selanjutnya. Evaluasi kinerja karyawan dilihat dari aspek-aspek yang telah ditentukan oleh perusahaan dimana karyawan yang dievaluasi akan mempunyai nilai evaluasi yang berbeda-beda untuk setiap

aspeknya. Dengan nilai evaluasi yang berbeda-beda tersebut akan menjadi suatu masalah apabila dikaitkan dengan hasil keputusan yang akan berguna untuk mengembangkan sumber daya manusia dalam suatu perusahaan.

Salah satu proses yang dapat digunakan untuk mengevaluasi kinerja karyawan dengan melakukan proses *data mining* yaitu klasifikasi[8]. Metode pengklasifikasian yang digunakan dalam penelitian ini ialah *K-Nearest Neighbor* (KNN). Algoritma *K-Nearest Neighbor* (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut[10].

Metode *Fuzzy K-Nearest Neighbor* merupakan metode klasifikasi yang menggabungkan teknik *fuzzy* dan *K-Nearest Neighbor Classifier*. Pada penelitian terdahulu menunjukkan bahwa algoritma *Fuzzy K-Nearest Neighbor* didapat dari hasil alokasi keanggotaan kelas *fuzzy* untuk vektor masukan, sehingga menghasilkan keputusan *fuzzy* berdasarkan label *fuzzy*[1]. Algoritma *Fuzzy K-Nearest Neighbor* dikembangkan memanfaatkan keanggotaan kelas *fuzzy* set sampel sehingga menghasilkan aturan klasifikasi *fuzzy*. Hasil dari kedua *crisp* dan aturan *Fuzzy KNN* dibandingkan pada dua *dataset*, dan algoritma *fuzzy* terbukti mendominasi *crisp* dengan memiliki tingkat kesalahan lebih rendah dan dengan memproduksi nilai-nilai keanggotaan yang berfungsi sebagai alat ukur dalam klasifikasi..

II. TINJAUAN PUSTAKA

2.1. Sumber Daya Manusia

Sumber daya manusia dalam sebuah organisasi perlu dikelola secara profesional agar terwujud keseimbangan antara kebutuhan pegawai dengan tuntutan dan kemampuan organisasi perusahaan. Keseimbangan tersebut merupakan kunci utama perusahaan agar dapat berkembang secara produktif dan wajar[6].

2.2. Evaluasi Kinerja

Evaluasi kinerja merupakan suatu proses penilaian kinerja aparatur yang dilakukan untuk melihat tanggung jawab pekerjaannya setiap hari apakah terjadi peningkatan atau penurunan sehingga pemimpin bisa memberikan suatu motivasi penunjang untuk melihat kinerja aparatur kedepannya. Evaluasi harus sering dilakukan agar masalah

yang di hadapi dapat diketahui dan dicari jalan keluar yang baik[7].

2.3. Logika Fuzzy

Logika *fuzzy* adalah suatu cara yang tepat untuk memetakan ruang *input* kedalam suatu ruang *output*[4]. Logika *fuzzy* bekerja dengan menggunakan derajat keanggotaan dari sebuah nilai yang kemudian digunakan untuk menentukan hasil yang ingin dihasilkan berdasarkan atas spesifikasi yang telah ditentukan.

2.4. Data Mining

Data mining adalah suatu algoritma di dalam menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*) yang sangat besar sehingga ditemukan satu pola yang menarik yang sebelumnya tidak diketahui.

2.5. Proses Data Mining

Prosedur yang umum digunakan untuk permasalahan data mining meliputi tahap-tahap sebagai berikut[3]:

1. Menentukan permasalahan dan merumuskan hipotesis

Pada tahap ini, ditentukan variabel-variabel dan hipotesis awal.

2. Mengumpulkan data

Tahap ini berkaitan dengan bagaimana data dihasilkan dan dikumpulkan.

3. *Preprocessing* data

Dilakukan pembersihan terhadap *outlier*, penanganan *missing value* maupun transformasi data.

4. Memperkirakan model

Pemilihan dan penerapan teknik data mining yang sesuai adalah tugas utama dalam tahap ini.

5. Menafsirkan model dan menarik kesimpulan.

Dalam skripsi digunakan normalisasi *min-max*. Normalisasi *min-max* dihitung dengan persamaan 1[8]. Normalisasi *min-max* memiliki keunggulan yaitu menjaga relasi pada data[2]. Serta mempunyai fungsi yaitu menyatukan satuan dari berbagai atribut.

$$V' = \frac{V - \min A}{\max A - \min A} \quad (1)$$

2.6. Algoritma K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data

pembelajaran yang jaraknya paling dekat dengan objek tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak *euclidean*.

2.7. Proses K-Nearest Neighbor (KNN)

Prinsip kerja *K-Nearest Neighbor (KNN)* adalah mencari jarak terdekat antara data yang dievaluasi dengan k tetangga terdekatnya dalam data pelatihan. Persamaan perhitungan untuk mencari euclidean ditunjukkan oleh persamaan 2.

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

2.8. Fuzzy K-Nearest Neighbor

Fuzzy K-NN merupakan metode klasifikasi yang menggabungkan teknik *fuzzy* dengan *k-Nearest Neighbor classifier*. Algoritma *fuzzy K-nearest neighbor* memberikan nilai keanggotaan kelas pada vektor sampel dan bukan menempatkan vektor pada kelas tertentu. FK-NN merupakan metode klasifikasi yang digunakan untuk memprediksi data uji menggunakan nilai derajat keanggotaan data uji pada setiap kelas. Kemudian diambil kelas dengan nilai derajat keanggotaan terbesar dari data uji sebagai kelas hasil prediksi.

Sebuah data memiliki nilai keanggotaan pada setiap kelas yang berbeda dengan nilai derajat keanggotaan dalam interval [0,1]. Teori himpunan fuzzy men-generalisasi teori K-NN klasik dengan mendefinisikan nilai keanggotaan sebuah data pada masing-masing kelas. Formula yang digunakan ditunjukkan oleh persamaan 3.

$$u_i(x) = \frac{\sum_{j=1}^k u_{ij} (\|x-x_j\|^{-2/(m-1)})}{\sum_{j=1}^k (\|x-x_j\|^{-2/(m-1)})} \quad (3)$$

Nilai u_{ij} pada $u_i(x)$ terlebih dahulu diproses dengan menggunakan persamaan 4.

$$u_{ij} = \begin{cases} 0.51 + (n_j/K) * 0.49, & \text{jika } j=i \\ (n_j/K) * 0.49, & \text{jika } j \neq i \end{cases} \quad (4)$$

2.9 Data Evaluasi Kinerja Karyawan PT Sierad Produce, Tbk

Data Evaluasi Kinerja Karyawan PT Sierad Produce, Tbk merupakan data hasil penilaian kinerja karyawan yang dilakukan oleh departemen *Human Resources Development (HRD)* PT Sierad Produce, Tbk. Data ini terdiri

dari 9 atribut, yaitu kooperatif, kehadiran dan ketepatan waktu, keandalan, inisiatif, sikap, judgement, komunikasi, hubungan antar manusia, dan keahlian profesional.

2.10 Perhitungan akurasi

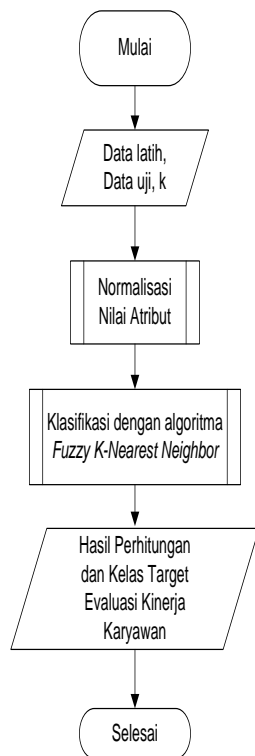
Akurasi merupakan seberapa dekat suatu angka hasil pengukuran terhadap angka sebenarnya (*true value atau reference value*). Dalam penelitian ini akurasi diagnosis dihitung dari jumlah diagnosis yang tepat dibagi dengan jumlah data. Tingkat akurasi diperoleh dengan perhitungan sesuai dengan persamaan 5[9].

$$\text{Akurasi} = \frac{\sum \text{data uji benar}}{\sum \text{jumlah total data uji}} \times 100\% \quad (5)$$

III. METODOLOGI DAN PERANCANGAN SISTEM

Secara umum sistem yang dibangun adalah suatu perangkat lunak untuk mengklasifikasi kinerja karyawan yang mengimplementasikan metode fuzzy k-nearest neighbor (FK-NN). Sistem ini bertujuan untuk menentukan apakah seorang karyawan mengikuti training/pelatihan atau tidak mengikuti training/pelatihan berdasarkan beberapa atribut yang digunakan. Terdapat beberapa tahap dalam sistem ini yang digunakan untuk melakukan klasifikasi menggunakan algoritma Fuzzy K-Nearest Neighbor.

Dalam proses sistem, diperlukan input berupa data evaluasi kinerja karyawan dan nilai k. Inputan tersebut dilakukan proses normalisasi dan klasifikasi dengan algoritma fuzzy k-nn. Alur sistem ini digambarkan dalam bentuk *flowchart* seperti pada gambar 1.



Gambar 1 Alur Proses Sistem

Proses normalisasi nilai atribut merupakan proses transformasi berupa normalisasi terhadap nilai-nilai atribut yang berguna untuk membakukan skala pengaruh yang ada pada atribut terhadap hasil. Langkah pertama yang dilakukan ialah memasukkan inputan berupa data latih dan data uji evaluasi kinerja karyawan, kemudian dicari nilai maksimum dan nilai minimum untuk tiap-tiap atribut. Setelah didapatkan nilai maksimum dan minimum dilakukan perhitungan normalisasi dan didapatkan hasil nilai atribut yang ternormalisasi.

Pada proses klasifikasi dengan algoritma *Fuzzy K-Nearest Neighbor* yang berguna untuk mengklasifikasikan hasil perhitungan jarak antara dua record menggunakan *Euclidean distance* dengan menggunakan inputan hasil nilai atribut yang ternormalisasi pada proses sebelumnya dan menghitung nilai derajat keanggotaan serta memilih nilai derajat keanggotaan yang terbesar untuk menentukan kelas target.

IV. HASIL DAN PEMBAHASAN

Pada pengujian klasifikasi digunakan dataset evaluasi karyawan yang terdiri dari 300 record. Data yang digunakan dalam sistem yaitu data evaluasi karyawan yang hasil evaluasi dibagi menjadi dua kelas. Terdapat 1 macam pengujian yang dilakukan, yaitu untuk mengetahui pengaruh nilai k dan data latih terhadap tingkat akurasi.

Pada pengujian ini terdapat dua parameter antara lain menentukan jumlah data latih dan jumlah nilai k. Pada pengujian jumlah data latih terdapat 6 macam data latih yang terdiri dari 80, 110, 140, 170, 200 dan 230 data dengan perulangan nilai $k=2$ sampai dengan $k=40$.

Pengujian data evaluasi kinerja karyawan ini didapatkan hasil pengujian dengan nilai akurasi maksimum pada jumlah 80 data latih adalah 54% dengan $k = 2$ dan nilai akurasi minimumnya tetap pada $k = 22$ sampai dengan $k = 40$ yaitu 43%.

Untuk 110 data latih didapatkan akurasi maksimumnya pada $k = 2$ yaitu 52% dan akurasinya mengalami penurunan dengan bertambahnya jumlah nilai k. Sedangkan akurasi minimum yang didapat ialah 43%.

Pada 140 data latih, akurasi maksimumnya ialah 53% yang didapat pada nilai $k=4$. Sedangkan untuk nilai akurasi minimumnya yaitu 44%.

Untuk pengujian 170 data latih, akurasinya mengalami penurunan dengan akurasi maksimumnya 52% dan akurasi minimumnya yaitu 44%.

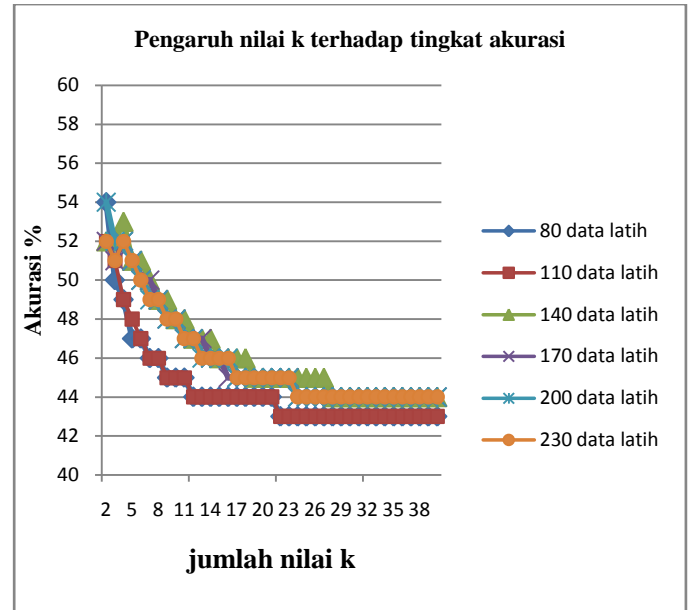
Pengujian data evaluasi kinerja karyawan untuk 200 data latih akurasinya mengalami penurunan, dengan akurasi maksimum yang didapat yaitu 54% dan hasil akurasi minimumnya tetap pada $k = 24$ sampai dengan $k = 40$ yaitu 44%.

Sedangkan untuk pengujian 230 data latih akurasi maksimumnya 52% dan akurasi minimumnya yaitu 44%. Hasil pengujian dataset evaluasi kinerja karyawan ditunjukkan pada tabel 1.

Tabel 1 Pengujian pada 80 – 230 data latih dengan nilai k = 2 – 70

80 data latih		110 data latih		140 data latih		170 data latih		200 data latih		230 data latih	
k	Akuisi %	k	Akuisi %	k	Akuisi %	k	Akuisi %	k	Akuisi %	k	Akuisi %
2	54	2	52	2	52	2	52	2	54	2	52
..
4	43	4	43	4	44	4	44	4	44	4	44
0		0		0		0		0		0	

Pada tahap pengujian ini didalamnya terdapat 6 macam data latih dengan k perulangan 2 sampai dengan 40. Untuk pengujian pengaruh nilai k sendiri dapat dilihat bahwa akurasi maksimum data evaluasi kinerja karyawan terdapat pada k=2. Tetapi pengaruh nilai k pada akurasi minimum tidak sama pada setiap record, sehingga dapat dilihat pada grafik bahwa nilai k minimumnya tidak dominan seperti nilai maksimum. Dilihat pada gambar 1, jumlah 80 data latih terletak pada grafik warna biru, 110 data latih pada grafik warna merah, 140 data latih pada grafik warna hijau, 170 data latih pada grafik warna ungu, 200 data latih pada grafik warna biru muda dan data latih dengan jumlah 230 record terletak pada grafik warna kuning, mengalami penurunan yang stabil pada tiap record Pengaruh nilai k terhadap tingkat akurasi ditunjukkan pada gambar 2.



Gambar 2 Pengaruh nilai k terhadap tingkat akurasi

Berdasarkan hasil pengujian data evaluasi kinerja karyawan bahwa akurasi paling baik pada sebagian besar data latih adalah saat nilai k = 2, yaitu 52% dan 54%. Akurasi tersebut tidak mencapai lebih dari 60%, karena jumlah record untuk tiap kelas tidak seimbang, dimana record dengan kelas 2 mendominasi dataset. Selain dilakukan pengujian dengan salah satu kelas yang dominan pada data, dilakukan juga pengujian terhadap data latih dengan komposisi jumlah kelas yang sama. Didapatkan hasil pengujian dengan akurasi mengalami kenaikan sampai 66%.

Gambar 2 menunjukkan bahwa tingkat akurasi dari k = 2 sampai dengan k = 40 grafiknya mengalami penurunan sampai akurasi minimumnya yaitu 43%. Dan pada grafik yang ditampilkan oleh gambar 2 ditunjukkan nilai akurasi yang stagnan pada nilai k = 22 sampai dengan k = 40 yaitu 43% untuk 80 dan 110 data latih, untuk 140 data latih mengalami akurasi yang stagnan pada nilai k = 28 sampai dengan k = 40 yaitu 44%. Sedangkan untuk 170 sampai dengan 230 data latih mengalami nilai akurasi yang stagnan pada nilai k = 24 sampai dengan k = 40 yaitu sebesar 44%.

Hasil pengujian yang didapat untuk mengetahui pengaruh nilai k (tetangga terdekat) terhadap tingkat akurasi yaitu, semakin bertambahnya jumlah nilai k (tetangga terdekat) maka akurasi yang didapatkan semakin menurun dan stabil pada nilai k (tetangga terdekat) tertentu untuk

masing masing data latih. Hal ini dikarenakan oleh nilai membership yang digunakan untuk penentuan kelas target pada penelitian bernilai maksimum. Kemudian akurasi k (tetangga terdekat) tertinggi cenderung pada $k = 2$, hal ini dikarenakan tetangga yang dibandingkan hanya dua tetangga saja. Untuk data yang berdekatan lebih sering memiliki kelas yang sama, sehingga mempengaruhi perhitungan membership (nilai derajat keanggotaan).

V. KESIMPULAN

Dari hasil uji dan analisa yang telah dilakukan, dapat diambil kesimpulan diantaranya ialah:

1. Algoritma Fuzzy K-Nearest Neighbor (FK-NN) dapat diterapkan pada dataset evaluasi kinerja karyawan. Terdapat dua proses utama yang dilakukan yaitu proses pertama ialah proses normalisasi nilai atribut pada dataset dan proses selanjutnya yaitu mengklasifikasikan hasil normalisasi nilai atribut dengan algoritma fuzzy k-nearest neighbor yang dimana didalamnya terdapat proses yang menghasilkan nilai derajat keanggotaan. Dengan dipilihnya nilai derajat keanggotaan yang terbesar yang digunakan untuk menjadi penentu kelas target yang baru.
2. Tingkat akurasi pada metode Fuzzy K-NN dipengaruhi oleh nilai tetangga terdekat. Nilai tetangga terdekat yang terlalu tinggi menghasilkan akurasi yang baik dimana nilai akurasi terbesarnya mencapai 66% dengan jumlah komposisi kelas yang sama pada data latih, yang disebabkan oleh nilai membership yang digunakan untuk penentuan kelas target pada penelitian bernilai maksimum.

DAFTAR PUSTAKA

- [1] Bezdek, J.C. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithm*. Plenum Press. New York.
- [2] Jayalakshmi, T. dan Santhakumaran, A. 2011. *Statistical Normalization and Backpropagation for Classification (Online)*, (<http://www.ijcte.org/papers/288-L052.pdf>, diakses tanggal 20 Mei 2012)
- [3] Kantardzic, Mehmed. 2003. *Data Mining : Concepts, Models, Methods and Algorithm*. John Wiley & Sons. New York.
- [4] Kusumadewi, S dan Purnomo, H. 2003. *Artificial Intelligence (Teknik & Aplikasinya)*. Graha Ilmu. Jogjakarta.
- [5] Mangkunegara, Anwar Prabu. 2000. *Manajemen Sumber Daya Manusia*. Penerbit Remaja Rosdakarya. Bandung.
- [6] Mangkunegara, Anwar Prabu. 2004. *Manajemen Sumber Daya Manusia Perusahaan*. PT. Remaja Rosdakarya. Bandung.
- [7] Mangkunegara, Anwar Prabu. 2005. *Evaluasi Kinerja SDM*. Penerbit PT. Refika Aditama. Bandung.
- [8] Moradian, Mehdi dan Baarani, Ahmad. 2009. *KNNBA: k-Nearest-Neighbor-Based-Association Algorithm (Online)*, (<http://www.jatit.org/volumes/research-papers/Vol6No1/14Vol6No1.pdf>, diakses tanggal 20 Mei 2012).
- [9] Nugraha, Dany, dkk. 2006. *Diagnosis Gangguan Sistem Urinari pada Anjing dan Kucing Menggunakan VFI 5*. Institut Pertanian Bogor.
- [10] Rakhman, Juniar P., Ramadijanti, Nana., Satriyanto, Edi. 2007. *Translasi Bahasa Isyarat*. Jurusan Teknik Informatika, PENS-ITS. Surabaya.
- [11] Wibowo, Luqman Putra. 2011. *Rancang Bangun Human Resource Information System (HRIS) Sebagai Aplikasi Penilaian Kinerja Karyawan Perusahaan Asuransi*. Institut Teknologi Negeri Surabaya. Surabaya.