

KLASIFIKASI PASIEN *SUSPECT PARVO* DAN *DISTEMPER* PADA DATA REKAM MEDIK RUMAH SAKIT HEWAN IPB MENGUNAKAN *VOTING FEATURE INTERVALS*

Muhammad Iqbal¹, Aziz Kustiyo¹, Ekowati Handharyani²

¹ Departemen Ilmu Komputer, Fakultas Matematika dan IPA, Institut Pertanian Bogor

² Departemen Klinik Patologi dan Reproduksi, Fakultas Kedokteran Hewan,
Institut Pertanian Bogor

ABSTRAK

Diagnosis terhadap suatu penyakit tertentu sudah merupakan kegiatan rutin yang dilakukan di dalam dunia kedokteran. Hal tersebut tidaklah mudah, karena banyak dari gejala-gejala penyakit yang ada merujuk pada beberapa penyakit tertentu. Penelitian ini bertujuan untuk membandingkan diagnosis tahap awal pada anjing yang dilakukan oleh dokter dengan diagnosis yang dihasilkan dari klasifikasi menggunakan voting feature intervals, dan untuk mengetahui tingkat akurasi yang dicapai oleh algoritma klasifikasi voting feature intervals (VF15).

Terdapat 2 proses dalam penelitian yaitu proses pelatihan dan pengujian algoritma. Hasil pelatihan menunjukkan bahwa terdapat gejala-gejala yang konsisten menjadi ciri khas kelas parvo saja atau kelas distemper saja pada setiap iterasi. Untuk kelas parvo terdapat 14 gejala yang menjadi ciri khas, sedangkan untuk kelas distemper terdapat 11 gejala yang menjadi ciri khas. Hasil pengujian juga menunjukkan terdapat 3 instance yang kelas prediksinya berdasarkan hasil klasifikasi menggunakan algoritma VF15 tidak sesuai dengan kelas sebenarnya pada data. Hal tersebut terjadi karena ketiga instance memiliki ciri khas gejala kelas prediksi.

Kata Kunci: *voting feature intervals, differential diagnosis, parvo, distemper.*

PENDAHULUAN

Latar Belakang

Diagnosis terhadap suatu penyakit tertentu merupakan kegiatan yang rutin dilakukan di dalam dunia kedokteran. Diagnosis ini merupakan hal yang sangat menentukan dalam memberikan tindakan perawatan selanjutnya yang dibutuhkan oleh pasien. Pada saat dilakukan diagnosis, dokter harus memperhatikan dengan jelas gejala-gejala yang dialami oleh pasien. Hal tersebut tidaklah mudah, karena banyak dari gejala-gejala penyakit yang ada merujuk pada beberapa penyakit tertentu maka dari itu dibutuhkan ketelitian, ketepatan, dan kecermatan dalam melakukan diagnosis sehingga kesalahan diagnosis suatu penyakit dapat dihindari.

Kondisi ini tidak hanya berlaku pada diagnosis terhadap manusia. Diagnosis pada hewan yang sakit juga harus dilakukan secara teliti, tepat, dan cermat. Diagnosis pada hewan tidaklah mudah bahkan dapat dikatakan diagnosis pada hewan lebih sulit daripada diagnosis pada manusia karena hewan tidak dapat memberitahukan keluhan yang dialaminya pada saat hewan itu sakit.

Diagnosis yang tepat membutuhkan perhatian ekstra dari para dokter hewan.

Diagnosis suatu penyakit tertentu dilakukan berdasarkan temuan-temuan klinis yang terdapat pada pasien. Temuan klinis ini diperoleh dari hasil pemeriksaan yang dilakukan oleh dokter. Temuan klinis ini meliputi suhu tubuh, berat badan serta gejala-gejala yang terdapat pada pasien. Diagnosis yang dilakukan pada tahap awal masih dikelompokkan ke dalam kategori *suspect* karena untuk keakuratan yang pasti harus dilakukan pemeriksaan laboratorium jika kategori *suspect* yang dimaksud termasuk ke dalam penyakit yang disebabkan oleh virus.

Kasus penyakit *parvo* yang menyerang anjing banyak ditemukan selama ini. *Parvo* merupakan penyakit yang disebabkan oleh virus yang muncul pertama kali pada tahun 1978. Karena kehebatan penyakit ini yang secara cepat menyebar melalui populasi anjing, *parvo* telah menarik perhatian publik. Penyakit ini sangat berbahaya terutama bagi anak anjing, karena anak anjing bisa menjadi sangat lemah dan sakit pada tahap awal dan juga dapat mengalami kematian mendadak tanpa adanya tanda-tanda sakit pada saluran pencernaan hanya

setelah periode yang singkat saat anjing mulai mengalami sakit. Virus yang menyebabkan penyakit ini serupa dengan virus yang menyebabkan penyakit *distemper* pada anjing, sehingga gejala yang ditimbulkan kedua penyakit tersebut biasanya hampir sama, ini menyebabkan *distemper* termasuk dalam salah satu penyakit yang dibandingkan dengan *parvo* dalam *differential diagnosis*.

Tujuan

Tujuan penelitian ini adalah untuk membandingkan diagnosis tahap awal pada *canine* (anjing) yang dilakukan oleh dokter dengan diagnosis yang dihasilkan dari klasifikasi menggunakan *voting feature intervals*. Penelitian ini juga bertujuan untuk mengetahui tingkat akurasi yang dicapai oleh algoritma klasifikasi *voting feature intervals* (VFI5) yang diterapkan pada data rekam medik rumah sakit hewan IPB.

Ruang Lingkup

Ruang lingkup pada penelitian ini meliputi :

1. Penerapan algoritma klasifikasi *voting feature intervals* (VFI5) pada data rekam medik pasien rumah sakit hewan IPB.
2. Klasifikasi dilakukan berdasarkan data yang terdapat pada atribut temuan klinis.

Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat pada bidang kedokteran hewan. Klasifikasi yang dilakukan dengan algoritma VFI5 diharapkan dapat dijadikan pembanding oleh dokter hewan.

TINJAUAN PUSTAKA

Parvo

Parvo merupakan penyakit pada anjing yang disebabkan oleh virus yang dinamakan *Canine Parvovirus* atau *Parvoviral Enteritis*. Virus tersebut tumbuh di dalam pembelahan sel yang cepat (Klinkam 1999). Saluran pencernaan pada anak anjing mempunyai konsentrasi terbesar untuk pembelahan sel secara cepat, sehingga penyakit ini lebih sering menyerang anak anjing daripada anjing dewasa. Pada beberapa kasus virus ini juga dapat menginfeksi otot jantung yang akhirnya mengarah kepada kematian mendadak (Klinkam 1999).

Parvo termasuk jenis penyakit sistemik yang akut yang biasanya ditandai dengan pendarahan pada radang usus (Tilley & Smith 1997). Infeksi penyakit ini sudah dimulai saat anjing memiliki gejala yang berhubungan dengan masalah pencernaan, khususnya bila terdapat gejala lesu, muntah-muntah, dan keluarnya kotoran dengan cairan atau darah yang berlebihan.

Pada temuan pemeriksaan klinis biasanya terdapat gejala-gejala seperti berikut :

- Demam
- Lesu
- Depresi
- Nafsu makan berkurang

Pada tahap lanjut gejala yang muncul dapat berupa :

- Diare (diare berdarah)
- Muntah-muntah
- Dehidrasi

Distemper

Distemper merupakan penyakit pada anjing yang disebabkan oleh virus yang dinamakan *Canine Distemper Virus*. Virus ini memperbanyak diri di dalam kelenjar getah bening dan pada akhirnya penyebaran virus akan sampai pada permukaan *epithelium* (jaringan penutup permukaan dalam alat-alat tubuh yang berongga) saluran pernapasan, saluran pencernaan, saluran kemih-kelamin, dan sistem saraf pusat dimana ini memulai kerusakan yang menyebabkan timbulnya gejala. Seperti halnya dengan *parvo*, *distemper* lebih sering menyerang anak anjing daripada anjing dewasa.

Distemper termasuk jenis penyakit akut yang penyebarannya dapat melalui udara dan cairan tubuh dari hewan yang terinfeksi, sehingga kontak dengan hewan yang terinfeksi dapat menularkan virus penyakit ini (Tilley & Smith 1997). Infeksi penyakit ini sudah dimulai saat anjing mengalami demam. Demam ini biasanya disertai dengan keluarnya cairan dari mata dan hidung.

Pada temuan pemeriksaan klinis biasanya terdapat gejala-gejala seperti berikut :

- Demam
- Nafsu makan berkurang
- Depresi

Pada tahap lanjut gejala yang muncul dapat berupa :

- Diare
- *Pneumonia* (radang paru)
- *Rhinitis* (radang selaput lendir hidung)

- Muntah-muntah

Differential Diagnosis

Differential diagnosis adalah diagnosis yang dilakukan dengan membandingkan tanda-tanda klinis suatu penyakit dengan tanda-tanda klinis penyakit lain (Muda 2003). *Differential diagnosis* dilakukan berdasarkan pada pemeriksaan kondisi klinis yang spesifik dan hasil pemeriksaan tambahan seperti tes darah, biopsi otot, dan lain-lain.

Klasifikasi

Klasifikasi adalah proses menemukan sebuah himpunan model (atau fungsi) yang menggambarkan dan membedakan kelas-kelas data atau berbagai konsep. Tujuannya adalah untuk meramalkan kelas dari objek-objek yang label kelasnya belum diketahui (Han & Kamber 2001).

Proses menemukan sebuah model dilakukan pada sebuah himpunan data *training* (data pelatihan). Untuk proses klasifikasi (prediksi) pada data *testing* (data pengujian) dilakukan berdasarkan model yang diturunkan dari data pelatihan tersebut.

Voting Feature Intervals (VFI5)

Algoritma *voting feature intervals* (VFI5) adalah sebuah algoritma klasifikasi. Sebuah konsep diwakilkan dengan sebuah himpunan selang-selang fitur atau atribut pada setiap dimensi fitur secara terpisah (Demiröz & Güvenir 1997).

VFI5 merupakan algoritma klasifikasi *non-incremental* karena semua objek pada data pelatihan diproses sekali (Güvenir & Sirin 1996 diacu dalam Güvenir & Emeksiz 2000). Klasifikasi pada VFI berdasarkan *feature vote*. Semua fitur yang ada berpartisipasi dengan memberikan nilainya pada *voting* di antara kelas-kelas. Kelas yang menerima *vote* tertinggi ditentukan sebagai kelas yang diramalkan.

Selang setiap fitur dihasilkan algoritma VFI5 dari data pelatihan. Sebuah selang fitur dapat mewakili objek-objek (*instances*) dari himpunan beberapa kelas daripada sebuah kelas tunggal. Selang yang dihasilkan dari data pelatihan dapat berupa *range interval* dan *point interval*. Sebuah *range interval* didefinisikan sebagai sebuah himpunan nilai-nilai yang berurutan dari fitur yang diberikan. Sebuah *point interval* didefinisikan untuk sebuah nilai fitur tunggal, dimana hanya sebuah nilai tunggal

yang digunakan untuk mendefinisikan selang tersebut.

Algoritma VFI5 mampu menangani nilai yang tidak diketahui atau nilai yang hilang. Jika terdapat nilai yang tidak diketahui atau nilai yang hilang pada sebuah fitur, fitur tersebut memberikan nilai 0 untuk *vote* pada setiap kelas. Karena itu, fitur yang mengandung nilai yang hilang atau nilai yang tidak diketahui diabaikan. Mengabaikan fitur tersebut merupakan pendekatan yang sangat wajar dan masuk akal (Demiröz & Güvenir 1997).

Pada sebuah himpunan data tertentu, jika ditambahkan fitur yang tidak relevan tidak akan mempengaruhi akurasi dari algoritma VFI5. Algoritma VFI5 juga mampu menangani adanya fitur-fitur yang tidak relevan. Ini disebabkan adanya mekanisme *voting* yang digunakan dalam klasifikasi, dimana *vote* dari sebuah fitur yang tidak relevan sama untuk semua kelas sehingga hal ini tidak berpengaruh terhadap hasil keluarannya (Güvenir 1998).

Terdapat dua proses atau fase pada algoritma klasifikasi VFI5, yaitu proses pelatihan (*training*) dan proses klasifikasi (prediksi). Proses pelatihan bertujuan untuk menemukan sebuah model yang akan digunakan dalam proses klasifikasi.

1 Pelatihan

Pada proses ini akan dihasilkan selang-selang untuk setiap fitur yang ada. Sebuah selang mewakili himpunan nilai-nilai dari fitur yang diberikan.

Untuk menghasilkan selang fitur tertentu perlu diketahui *end point* atau batas-batas pada selang itu. Proses menemukan *end point* dibedakan untuk fitur linear dan fitur nominal. *End point* fitur linear, yaitu fitur dimana nilai-nilainya memiliki urutan dan dapat dibandingkan tingkatannya ditentukan dengan mencari nilai terbesar dan terkecil pada fitur tersebut untuk setiap kelas. Di lain pihak, *end point* fitur nominal, yaitu fitur dimana nilai-nilainya tidak memiliki urutan dan tidak dapat dibandingkan tingkatannya ditentukan dengan mencatat semua nilai yang berbeda pada fitur tersebut. Untuk fitur linear selang yang dihasilkan dapat berupa *point interval* dan *range interval* serta jumlah maksimum *end point* dan selang yang dihasilkan adalah $2k$ dan $4k+1$ dimana k adalah jumlah kelas, sedangkan untuk fitur nominal selang yang dihasilkan hanya berupa *point interval*.

Untuk setiap selang i dari sebuah fitur f dihitung jumlah *instance* pelatihan setiap kelas c yang jatuh pada selang i dan hasilnya disimpan sebagai $interval_class_count[f,i,c]$. Hasil proses ini merupakan *vote* kelas c pada selang i .

Jumlah *instance* untuk setiap kelas c dapat berbeda-beda, sehingga untuk menghilangkan efek perbedaan distribusi setiap kelas, *vote* kelas c untuk fitur f pada selang i dinormalisasi. Normalisasi dilakukan dengan membagi jumlah *instance* pelatihan setiap kelas c yang jatuh pada selang i sebuah fitur f dengan jumlah *instance* setiap kelas c dan hasilnya disimpan sebagai $interval_class_vote[f,i,c]$. Kemudian nilai-nilai $interval_class_vote[f,i,c]$ dinormalisasi kembali sehingga jumlah *vote* setiap kelas c pada selang i untuk suatu fitur f sama dengan 1. Normalisasi bertujuan agar jumlah *instance* pelatihan setiap kelas c tidak mempengaruhi *voting*, sehingga setiap fitur memiliki kekuatan *voting* yang sama. *Pseudocode* algoritma pelatihan VF15 disajikan pada Gambar 1.

2 Klasifikasi

Vote pada setiap kelas c diberi nilai awal sama dengan 0, karena semua fitur pada awalnya belum memberikan *vote*, kemudian dicari selang i dimana *instance* pengujian jatuh pada selang tersebut untuk setiap fitur f . Jika terdapat nilai suatu fitur dari *instance* pengujian yang hilang atau tidak diketahui, maka fitur tersebut diasumsikan tidak memberikan *vote* sehingga nilai *vote* untuk fitur tersebut sama dengan 0.

Setelah selang i dimana *instance* pengujian jatuh diketahui, *vote-vote* setiap kelas c pada selang tersebut disimpan dalam sebuah vektor $\langle feature_vote[f,C_1], \dots, feature_vote[f,C_j], \dots, feature_vote[f,C_k] \rangle$, dimana $feature_vote[f,C_j]$ merupakan *vote* fitur f untuk kelas C_j dan k adalah jumlah kelas. Kemudian nilai-nilai *vote* dari setiap fitur pada selang i dimana *instance* pengujian jatuh dijumlahkan setelah masing-masing dikalikan dengan bobot fitur yang bersesuaian dan hasilnya disimpan dalam sebuah vektor *vote* $\langle vote[C_1], \dots, vote[C_k] \rangle$. Kelas dengan jumlah *vote* terbesar diramalkan sebagai kelas dari *instance* pengujian. *Pseudocode* algoritma pengujian VF15 disajikan pada Gambar 2.

METODE PENELITIAN

Penelitian ini menerapkan algoritma klasifikasi VF15 pada data rekam medik pasien rumah sakit hewan IPB. Terdapat beberapa proses dalam penelitian ini. Tahapan pertama yang dilakukan adalah pencarian data, kemudian data yang telah didapatkan dipelajari dan dilihat karakteristiknya untuk selanjutnya digunakan dalam pelatihan dan pengujian.

Data

Data yang digunakan dalam penelitian ini adalah data rekam medik pasien jenis *canine* (anjing) rumah sakit hewan IPB, khususnya data rekam medik pasien *suspect parvo* dan *distemper*. Data rekam medik pasien rumah sakit hewan IPB memiliki atribut, yaitu :

- 1 Nomor rekam medik
Nomor yang terdaftar sesuai dengan administrasi dari manajemen rumah sakit.
- 2 *Signalement*
Keterangan individual dari pasien yang bersangkutan.
- 3 *Anamnesis*
Sejarah informasi penyakit yang pernah diderita oleh pasien yang bersangkutan atau dapat berupa keluhan yang disampaikan oleh pemilik hewan.
- 4 Temuan klinis
Temuan atau informasi yang diperoleh dari hasil pemeriksaan klinis.

Praproses

Pada tahap ini dilakukan pemilihan data pada data rekam medik. Pemilihan yang dilakukan yaitu pemilihan data dari atribut tertentu. Data yang dipilih merupakan data yang cukup relevan untuk dijadikan fitur-fitur.

Data Pelatihan dan Data Pengujian

Seluruh data yang digunakan dibagi secara acak menjadi beberapa himpunan bagian. Masing-masing himpunan bagian memiliki ukuran yang hampir sama. Data yang telah terbagi menjadi beberapa himpunan bagian tersebut digunakan dalam pelatihan dan pengujian. semua node pada jaringan per satu detik.

Algoritma VF15

Algoritma yang digunakan dalam penelitian ini adalah algoritma klasifikasi VF15. Algoritma ini memiliki dua proses atau fase,

```

train (TrainingSet):
begin
  for each feature f
    if f is linear
      for each class c
        EndPoints[f] = EndPoints[f] U find_end_points(TrainingSet, f, c);
      sort(EndPoints[f]);

      for each end point p in EndPoints[f]
        form a point interval from end point p
        form a range interval between p and the next endpoint ≠ p

    else /* f is nominal */
      form a point interval for each value of f

  for each interval i on feature f
    for each class c
      interval_class_count[f, i, c] = count_instance(f, i, c);
  for each interval i on feature f
    for each class c
      interval_class_vote[f, i, c] = interval_class_count[f, i, c] / class_count[c]
      normalize interval_class_vote[f, i, c];
      /*such that  $\sum_c \text{interval\_class\_count}[f, i, c] = 1$ */
end.

```

Gambar 1 Algoritma pelatihan VF15

```

classify(e):
/*e is example to be classified*/
begin
  for each class c
    vote[c] = 0 /*sum of vote of class c*/
  for each feature f
    for each class c
      feature_vote[f, c] = 0 /*vote of feature f for class c*/
    if ef value is known
      i = find_interval(f, ef)
      feature_vote[f, c] = interval_class_vote[f, i, c]
    for each class c
      vote[c] = vote[c] + (feature_vote[f, c] * w[f]);
  return class c with highest vote[c];
end.

```

Gambar 2 Algoritma pengujian VF15

yaitu proses pelatihan dan proses klasifikasi. Proses pelatihan menghasilkan sebuah model yang diturunkan dari data pelatihan. Model ini akan digunakan pada data pengujian dalam proses pengujian.

Pelatihan

Data pelatihan digunakan sebagai input dari algoritma VF15 pada proses pelatihan. Untuk setiap fitur akan dihasilkan selang-selang, kemudian pada akhirnya didapatkan nilai *vote* untuk setiap kelas pada selang-selang tersebut.

Pengujian

Untuk setiap fitur dari setiap *instance* pengujian, dicari selang dimana nilai fitur

dari *instance* pengujian itu jatuh. Kemudian dilihat *vote* setiap kelas pada selang tersebut. Selanjutnya nilai-nilai *vote* itu dijumlahkan sehingga kelas yang memiliki nilai *vote* terbesar adalah kelas prediksi dari *instance* pengujian tersebut.

Akurasi

Pada penelitian ini ingin diketahui tingkat akurasi yang dicapai algoritma VF15. Tingkat akurasi dihitung dengan cara :

$$\text{tingkat akurasi} = \frac{\sum \text{data uji benar diklasifikasi}}{\sum \text{total data uji}}$$

Spesifikasi Implementasi

Implementasi dirancang dan dibangun dengan perangkat keras dan perangkat lunak sebagai berikut :

Perangkat keras :

- 1 Prosesor AMD Sempron 2200+ 1.5GHz
- 2 Memori 256 MB
- 3 Harddisk 40 GB
- 3 Monitor 15"
- 4 Alat input mouse dan keyboard

Perangkat Lunak :

- 1 Microsoft® Windows XP Professional 2002 SP2
- 2 Microsoft® Internet Explorer 6.0
- 3 PHP 5.0.3
- 4 Apache Webserver

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data rekam medik pasien jenis *canine* (anjing) rumah sakit hewan IPB, khususnya data rekam medik pasien *suspect parvo* dan *distemper*. Data rekam medik yang berhasil dikumpulkan yaitu data tahun 2002 sampai dengan tahun 2006 sebanyak 30 *record*.

Semua *instance* yang memiliki nilai fitur yang tidak lengkap tetap digunakan karena VF15 memiliki mekanisme *voting* yang mampu menangani nilai fitur yang tidak diketahui. Dengan demikian, tidak ada pengurangan jumlah *instance* dari data awal dan *instance-instance* tersebut digunakan sebagai data pelatihan maupun data pengujian.

Pada data rekam medik terdapat beberapa atribut, salah satu di antaranya yaitu atribut temuan klinis. Data yang terdapat pada atribut temuan klinislah yang digunakan untuk diolah. Atribut temuan klinis pada data rekam medik yang digunakan mengandung informasi berupa temuan gejala-gejala yang diderita pasien. Gejala-gejala ini ditemukan berdasarkan hasil pemeriksaan klinis yang dilakukan oleh dokter.

Setiap gejala berbeda yang ditemukan diuraikan, sehingga terdapat 47 gejala dan 2 keterangan tambahan berupa suhu tubuh dan berat badan. Gejala-gejala dan 2 keterangan tambahan itu dijadikan sebagai fitur sehingga pada akhirnya terdapat fitur sebanyak 49 buah. Empat puluh sembilan fitur yang ada masing-masing dilambangkan

berturut-turut dengan variabel X_1, X_2, \dots, X_{49} . Spesifikasi data yang digunakan dalam penelitian disajikan pada Tabel 1.

Tabel 1 Spesifikasi data yang digunakan

Nama data	Jumlah <i>instance</i>	Jumlah fitur	Jumlah kelas
Rekam medik	30	49	2

Fitur-fitur yang ada dibedakan menjadi fitur linear dan fitur nominal. Keterangan tambahan berupa suhu tubuh dan berat badan merupakan fitur linear sedangkan 47 gejala berbeda yang ditemukan pada data merupakan fitur nominal. Nilai untuk fitur nominal ditentukan sebagai berikut :

- Nilai 1 diberikan untuk fitur nominal tertentu pada *instance* yang memiliki gejala penyakit yang dilambangkan fitur nominal tersebut.
- Nilai 0 diberikan untuk fitur nominal tertentu pada *instance* yang tidak memiliki gejala penyakit yang dilambangkan fitur nominal tersebut.

Data keseluruhan sebanyak 30 *instance* tersebut terlebih dahulu dibagi secara acak menjadi 3 himpunan bagian yang ukurannya hampir sama satu sama lain. Pembagian data keseluruhan secara acak menghasilkan himpunan bagian yang disebut sebagai himpunan bagian S_1 , himpunan bagian S_2 , dan himpunan bagian S_3 . Pembagian data keseluruhan secara acak ini menghasilkan himpunan bagian-himpunan bagian yang masing-masing memiliki jumlah *instance* sebanyak 10 buah. Hasil pembagian data keseluruhan disajikan pada Tabel 2.

Tabel 2 Hasil pembagian data keseluruhan

Himpunan bagian	<i>Parvo</i>	<i>Distemper</i>
S_1	5 <i>instance</i>	5 <i>instance</i>
S_2	7 <i>instance</i>	3 <i>instance</i>
S_3	5 <i>instance</i>	5 <i>instance</i>
Total	17 <i>instance</i>	13 <i>instance</i>

Pada penelitian ini pelatihan dan pengujian data dilakukan sebanyak 3 kali. Susunan data yang digunakan sebagai data pelatihan dan data pengujian pada setiap iterasi disajikan pada Tabel 3. Untuk setiap iterasi, dalam hal ini berarti data yang digunakan sebagai pelatihan sebanyak 20 *instance* sedangkan data yang digunakan sebagai pengujian sebanyak 10 *instance*.

Tabel 3 Susunan data pelatihan dan data pengujian

Iterasi	Pelatihan	Pengujian
Iterasi pertama	S_2 & S_3	S_1
Iterasi kedua	S_1 & S_3	S_2
Iterasi ketiga	S_1 & S_2	S_3

Iterasi Pertama

Pada iterasi pertama, himpunan bagian S_2 dan himpunan bagian S_3 digunakan sebagai data pelatihan sedangkan himpunan bagian S_1 digunakan sebagai data pengujian. Komposisi jumlah *instance* per kelas pada data pelatihan dan data pengujian pada iterasi ini disajikan pada Tabel 4.

Tabel 4 Komposisi jumlah *instance* per kelas data pelatihan dan data pengujian iterasi pertama

Kelas	Pelatihan	Pengujian
<i>Parvo</i>	12 <i>instance</i>	5 <i>instance</i>
<i>Distemper</i>	8 <i>instance</i>	5 <i>instance</i>

Proses pelatihan pada iterasi ini menghasilkan selang-selang fitur. Setiap selang suatu fitur tertentu memiliki nilai-nilai yang didistribusikan fitur tersebut untuk kelas *parvo* dan kelas *distemper*.

Untuk fitur X3 sampai dengan fitur X50, dapat dilihat nilai-nilai distribusi fitur-fitur tersebut pada suatu selang yang dihasilkan oleh proses pelatihan yang mencerminkan kecenderungan fitur-fitur tersebut untuk menjadi ciri khas gejala dari kelas *parvo* atau kelas *distemper*.

Pengujian yang dilakukan pada iterasi ini sebagai klasifikasi pada data pengujian S_1 menghasilkan akurasi sebesar 100%. Prediksi kelas sebagai hasil klasifikasi yang dilakukan oleh algoritma VF15 sama dengan kelas sebenarnya pada data rekam medik untuk seluruh data pengujian S_1 .

Iterasi Kedua

Pada iterasi kedua, himpunan bagian S_1 dan himpunan bagian S_3 digunakan sebagai data pelatihan sedangkan himpunan bagian S_2 digunakan sebagai data pengujian. Komposisi jumlah *instance* per kelas pada data pelatihan dan data pengujian pada iterasi ini disajikan pada Tabel 5.

Tabel 5 Komposisi jumlah *instance* per kelas data pelatihan dan data pengujian iterasi kedua

Kelas	Pelatihan	Pengujian
<i>Parvo</i>	10 <i>instance</i>	7 <i>instance</i>
<i>Distemper</i>	10 <i>instance</i>	3 <i>instance</i>

Proses pelatihan pada iterasi ini menghasilkan selang-selang fitur. Seperti pada iterasi pertama, setiap selang suatu fitur tertentu memiliki nilai-nilai yang didistribusikan fitur tersebut untuk kelas *parvo* dan kelas *distemper*.

Seperti pada iterasi pertama, untuk fitur X3 sampai dengan fitur X50, dapat dilihat nilai-nilai distribusi fitur-fitur tersebut pada suatu selang yang dihasilkan oleh proses pelatihan yang mencerminkan kecenderungan fitur-fitur tersebut untuk menjadi ciri khas gejala dari kelas *parvo* atau kelas *distemper*.

Pengujian yang dilakukan pada iterasi ini sebagai klasifikasi pada data pengujian S_2 menghasilkan akurasi sebesar 100%. Prediksi kelas sebagai hasil klasifikasi yang dilakukan oleh algoritma VF15 sama dengan kelas sebenarnya pada data rekam medik untuk seluruh data pengujian S_2 .

Iterasi Ketiga

Pada iterasi ketiga, himpunan bagian S_1 dan himpunan bagian S_2 digunakan sebagai data pelatihan sedangkan himpunan bagian S_3 digunakan sebagai data pengujian. Komposisi jumlah *instance* per kelas pada data pelatihan dan data pengujian pada iterasi ini disajikan pada Tabel 6.

Tabel 6 Komposisi jumlah *instance* per kelas data pelatihan dan data pengujian iterasi ketiga

Kelas	Pelatihan	Pengujian
<i>Parvo</i>	12 <i>instance</i>	5 <i>instance</i>
<i>Distemper</i>	8 <i>instance</i>	5 <i>instance</i>

Proses pelatihan pada iterasi ini menghasilkan selang-selang fitur. Seperti pada iterasi pertama dan iterasi kedua, setiap selang suatu fitur tertentu memiliki nilai-nilai yang didistribusikan fitur tersebut untuk kelas *parvo* dan kelas *distemper*.

Seperti pada iterasi pertama dan iterasi kedua, untuk fitur X3 sampai dengan fitur X50, dapat dilihat nilai-nilai distribusi fitur-fitur tersebut pada suatu selang yang dihasilkan oleh proses pelatihan yang mencerminkan kecenderungan fitur-fitur tersebut untuk menjadi ciri khas gejala dari kelas *parvo* atau kelas *distemper*.

Pengujian yang dilakukan pada iterasi ini sebagai klasifikasi pada data pengujian S_3 menghasilkan akurasi sebesar 70%. Terdapat tiga *instance* pada data pengujian S_3 yang kelas prediksinya tidak sesuai dengan kelas

sebenarnya pada data rekam medik, yaitu *instance* dengan nomor rekam medik 182.10.03, 002.01.05, dan 076.02.06. *Instance* dengan nomor rekam medik 182.10.03 dan 002.01.05 diprediksi oleh algoritma VFI5 termasuk ke dalam kelas *parvo*, sedangkan pada data rekam medik kelas sebenarnya dari *instance* tersebut adalah kelas *distemper*. Hal ini terjadi karena *instance* tersebut memiliki beberapa gejala yang merupakan ciri khas dari kelas *parvo* berdasarkan proses pelatihan pada iterasi ini. Gejala-gejala tersebut adalah lemas, muntah, diare, diare berdarah, palpasi *abdominal* sakit, bulu kusam yang berturut-turut dilambangkan dengan variabel X5, X7, X8, X9, X15, dan X31. Fitur X5, X7, X8, X9, X15, dan X31 memberikan nilai *vote* yang lebih besar untuk kelas *parvo* sehingga kelas *parvo* memiliki total nilai *vote* yang lebih besar dibandingkan dengan kelas *distemper*. Normalisasi dua *instance* pengujian ini disajikan pada Tabel 7.

Tabel 7 Normalisasi *instance* pengujian
182.10.03 dan 002.01.05

No. rekam medik	<i>Parvo</i>	<i>Distemper</i>
182.10.03	0.53	0.47
002.01.05	0.51	0.49

Instance pengujian dengan nomor rekam medik 182.10.03 dan 002.01.05 memiliki nilai normalisasi yang mendekati 0.5. Ini berarti kedua *instance* tersebut mempunyai peluang yang hampir sama untuk setiap kelasnya.

Instance berikutnya yang kelas prediksinya tidak sesuai dengan kelas sebenarnya pada data rekam medik, yaitu *instance* dengan nomor rekam medik 076.02.06. *Instance* ini diprediksi oleh algoritma VFI5 termasuk ke dalam kelas *distemper*, sedangkan pada data rekam medik kelas sebenarnya dari *instance* tersebut adalah kelas *parvo*. Hal ini terjadi karena *instance* tersebut memiliki beberapa gejala yang merupakan ciri khas dari kelas *distemper* berdasarkan proses pelatihan pada iterasi ini. Gejala-gejala tersebut adalah *faeces* lembek, *alopecia* punggung, *alopecia abdomen* yang berturut-turut dilambangkan dengan variabel X29, X37, dan X38. Fitur X29, X37, dan X38 memberikan nilai *vote* yang lebih besar untuk kelas *distemper* sehingga kelas *distemper* memiliki total nilai *vote* yang lebih besar dibandingkan dengan

kelas *parvo*. Normalisasi dua *instance* pengujian ini disajikan pada Tabel 8.

Tabel 8 Normalisasi *instance* pengujian
076.02.06

No. rekam medik	<i>Parvo</i>	<i>Distemper</i>
076.02.06	0.49	0.51

Instance pengujian dengan nomor rekam medik 076.02.06 memiliki nilai normalisasi yang mendekati 0.5. Ini berarti *instance* tersebut mempunyai peluang yang hampir sama untuk setiap kelasnya.

Hasil Pelatihan dan Hasil Pengujian

Proses pelatihan yang dilakukan pada setiap iterasi menghasilkan selang-selang fitur. Selang-selang untuk setiap fiturnya mempunyai nilai *vote* untuk kelas *parvo* dan kelas *distemper*. Untuk setiap fitur yang merupakan gejala, terdapat sebuah selang yaitu *point interval* 1 dimana nilai *vote* pada selang ini mencerminkan kecenderungan fitur tersebut untuk menjadi ciri khas gejala dari kelas yang ada.

Pada setiap iterasi terdapat fitur-fitur yang konsisten menjadi ciri khas gejala kelas *parvo* saja atau kelas *distemper* saja. Fitur-fitur tersebut adalah :

1 Kelas *parvo*

- X5 = Lemas
- X7 = Muntah
- X9 = Diare berdarah
- X15 = Palpasi *abdominal* sakit
- X16 = Palpasi *abdominal* tegang
- X19 = Turgor kulit jelek
- X23 = *Mucosa* pucat
- X25 = LGL *poplitea* bengkak
- X26 = LGL *prescapularis* bengkak
- X27 = LGL *praefemoralis* bengkak
- X28 = *Perineal* kotor
- X39 = *Lethargy*
- X44 = Air liur kental
- X49 = Dehidrasi

2 Kelas *distemper*

- X10 = Batuk
- X11 = Bersin
- X14 = Auskultasi paru *rough*
- X17 = Palpasi *trachea* batuk
- X20 = *Discharge* nasal
- X22 = *Mucosa* rose
- X34 = Bintik merah bagian *ventral*
- X35 = Bintik merah kulit *abdomen*
- X37 = *Alopecia* punggung
- X38 = *Alopecia* abdomen
- X46 = Lepuh pada kulit

Proses pengujian pada setiap iterasi menghasilkan akurasi. Akurasi dari setiap iterasi pada proses pengujian disajikan pada Tabel 9.

Tabel 9 Akurasi dari setiap iterasi

Iterasi	Akurasi
Pertama	100%
Kedua	100%
Ketiga	70%
Rata-rata	90%
Standar deviasi	17.32%

KESIMPULAN DAN SARAN

Kesimpulan

Algoritma *voting feature intervals* (VFI5) digunakan untuk melakukan klasifikasi. Data yang digunakan adalah data rekam medik pasien jenis *canine* (anjing) rumah sakit hewan IPB, khususnya data rekam medik pasien *suspect parvo* dan *distemper*.

Proses pelatihan pada setiap iterasi menghasilkan selang-selang fitur. Nilai *voting* pada *point interval* 1 untuk semua fitur yang merupakan gejala mencerminkan kecenderungan fitur tersebut untuk menjadi ciri khas gejala dari kelas *parvo* atau kelas *distemper*.

Dari fitur-fitur gejala yang ada terdapat 25 fitur yang konsisten menjadi ciri khas gejala kelas *parvo* saja atau kelas *distemper* saja pada setiap iterasi. Fitur-fitur tersebut terdiri dari 14 fitur untuk ciri khas gejala kelas *parvo* dan 11 fitur untuk ciri khas gejala kelas *distemper*.

Pengujian yang dilakukan sebanyak 3 kali menunjukkan terdapat 3 *instance* yang klasifikasinya tidak sesuai yaitu *instance* dengan nomor rekam medik 182.10.03, 002.01.05, dan 076.02.06.

Akurasi dari klasifikasi yang dihasilkan oleh algoritma VFI5 cukup tinggi untuk setiap iterasinya. Iterasi pertama menghasilkan akurasi sebesar 100%, iterasi kedua menghasilkan akurasi sebesar 100%, dan iterasi ketiga menghasilkan akurasi sebesar 70%. Rata-rata akurasi yang dihasilkan oleh algoritma VFI5 adalah sebesar 90% dan standar deviasinya adalah sebesar 17.32%.

Saran

Penelitian ini menggunakan bobot fitur yang seragam yaitu satu. Hal ini masih dapat

dikembangkan lebih lanjut dengan menggunakan bobot yang berbeda untuk setiap fitur.

Untuk validasi silang pada data juga dapat dikembangkan dengan menggunakan *leave-one-out cross validation*. Hasil yang didapatkan dengan menggunakan metode tersebut dapat dibandingkan dengan data yang menggunakan *k-fold cross validation*. Untuk pengembangan selanjutnya data yang akan digunakan pada penelitian diharapkan memiliki jumlah *record* yang lebih banyak.

DAFTAR PUSTAKA

- Demiröz G, Güvenir HA. 1997. *Classification by Voting Feature Intervals*. <http://www.cs.ucf.edu/~ecl/papers/demiroz97classification.pdf>. [5 Mei 2006]
- Güvenir HA. 1998. *A Classification Learning Algorithm Robust to Irrelevant Features*. http://www.cs.bilkent.edu.tr/tech-report/_1998/BU-CEIS-9810.pdf. [27 Juli 2006]
- Güvenir HA, Emeksiz N. 2000. *An Expert System for the Differential Diagnosis of Erythematous-Squamous Disease. Expert System with Applications*, Vol. 18, No.1, (2000), hlm 43-49.
- Han J, Kamber M. 2001. *Data Mining Concepts & Techniques*. USA : Academic Press.
- Klinkam M. 1999. *Canine Parvo, Parvo Virus and Parvovirus Disease, Symptoms and Treatment*. <http://www.nwk9.com/parvovirus.htm> [25 Juli 2006]
- Muda A. 2003. *Kamus Lengkap Kedokteran*. Surabaya : Gitamedia Press.
- Sarle W. 2004. *What are cross-validation and bootstrapping?*. <http://www.faqs.org/faqs/ai-faq/neuralnets/part3/section-12.html>. [Juli 2006]
- Tilley LP, Smith FWK. 1997. *The 5 Minute Veterinary Consult, Canine and Feline*. Baltimore : Williams & Wilkins.