

## ANALISIS DAN IMPLEMENTASI DEEPS: INSTANCE-BASED CLASSIFICATION MENGGUNAKAN EMERGING PATTERN PADA STUDI KASUS TEKNIK KLASIFIKASI GEJALA FISIK PENYAKIT DEMAM TIFOID (TIPUS)

Ahsan Agung<sup>1</sup>, Shaufiah<sup>2</sup>, Imelda Ataina<sup>3</sup>

<sup>1</sup>Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

### Abstrak

Klasifikasi sebagai sebuah task pada data mining bertujuan memprediksi keanggotaan masing-masing instan pada data, untuk memperoleh informasi dari sekumpulan data.

Pengklasifikasian data gejala tipus pada tugas akhir ini menggunakan predictive method DeEPs (Decision making by Emerging Patterns). DeEPs merupakan metoda klasifikasi berbasis instan yang artinya setiap record pada data testing akan dibandingkan dengan data training untuk mendapatkan fungsi solusi lokal. Pencarian fungsi solusi lokal ini memanfaatkan konsep Emerging pattern. Emerging pattern merupakan itemset yang memiliki frekuensi berubah secara signifikan antara kelas pada suatu data. Khusus pada metoda klasifikasi DeEPs pola atau itemset yang diamati adalah pola yang relatif muncul hanya pada suatu kelas saja atau dikenal dengan JEP (Jumping Emerging Pattern). Setelah JEP didapatkan maka frekuensi kemunculannya dihitung dan dibandingkan untuk pengelompokan kelas.

Implementasi metoda DeEPs membutuhkan efisiensi untuk menekan jumlah powerset dari pola yang dicari saat operasi selisih antar maksimum representasi kelas. Hasil klasifikasi DeEPs terhadap data gejala tipus menunjukkan performansi yang baik dari segi akurasi dan waktu klasifikasi.

**Kata Kunci :** klasifikasi, predictive method, DeEPs (Decision making by Emerging Pattern), metoda klasifikasi berbasis instan, data testing, data training, fungsi solusi lokal, Emerging pattern, itemset, JEP (Jumping Emerging Pattern).

### Abstract

Classification as one of data mining task which purpose to predict the membership each instant at data, to gain information.

Classification typhus data at this task using predictive method, which is DeEPs (Decision making by Emerging Patterns). DeEPs is an instance-based classification method which mean each record at testing data will be compared with training data to get local solution function. Emerging Pattern concepts are used to get this local solution functions. Emerging Pattern is itemset or pattern where emergence frequency change significantly between classes at data. Particularly at DeEPs classification method the observed pattern are that only appear in a class, known as JEP (Jumping Emerging Pattern). After Jumping Emerging Patterns are founded then these JEPs frequency will be counted and compared to determine the class a record belong.

Implementation of DeEPs requires some efficiency to reduce number of powerset of patterns that are searched during difference operation among maximum representation of the class. From experiment which has done DeEPs show good result in terms of accuration and time of classification.

**Keywords :** predictive method, DeEPs (Decision making by Emerging Pattern), instance-based classification method, testing data, training data, local solution function, Emerging pattern, itemset, JEP (Jumping Emerging Pattern)

# 1 PENDAHULUAN

## 1.1 Latar belakang masalah

Klasifikasi sebagai sebuah *task* pada data mining bertujuan memprediksi keanggotaan masing-masing instan pada data untuk memperoleh informasi dari sekumpulan data. Dalam upaya mendapatkan kelas suatu instan dapat menerapkan dua pendekatan *eager learning* atau pun *lazy learning*.

Pada metode klasifikasi dengan pendekatan *eager learning* sebelum proses klasifikasi didahului dengan pembuatan model atau fungsi yang menjelaskan atau membedakan konsep kelas data melalui proses *learning*. Keunggulan proses *learning* pada metode *eager learning* walau membutuhkan waktu untuk mendapatkan fungsi global, namun lebih sedikit waktu dan ruang memori untuk mengklasifikasi dibanding *lazy learning*[11]. Kekurangan utama pendekatan *eager learning* secara umum tidak bisa memberikan fungsi solusi lokal atau *cased based reasoning*[11]. Sebaliknya pada pendekatan *lazy learning*, fungsi atau rule yang didapat proses *learning* bersifat lokal, namun hasilnya akan memiliki ketergantungan dengan data *training* yang ada, dengan asumsi semakin optimal jumlah instan, hasil klasifikasi akan semakin baik. Kekurangan pendekatan *lazy learning* adalah walau metoda *lazy learning* sederhana dan cukup efektif namun biasanya membutuhkan waktu dan memori yang lebih banyak dari *eager learning* untuk proses klasifikasi[12].

*Instance-based learning* merupakan macam dari *lazy learning* dimana proses *learning* dilakukan dengan membandingkan instan atau *record* pada data *testing* dengan masing-masing instan pada data *training* [14]. Contoh *instance-based learning* diantaranya adalah KNN dan variannya yang memanfaatkan jarak Euclidean untuk klasifikasi. Namun jarak Euclidean hanya digunakan untuk variabel kontinu. Untuk klasifikasi teks KNN bisa menggunakan jarak Hamming[13]. DeEPs (*Decision making by Emerging Patterns*) merupakan *instance based learning* namun dengan konsep yang berbeda dengan memanfaatkan frekuensi kemunculan pola (*Emerging Pattern*) untuk klasifikasi. DeEPs dimanfaatkan untuk menangani masalah prediksi yang membutuhkan pengetahuan akan pola dan rule yang komprehensif pada aplikasi yang cerdas seperti klasifikasi tipe sel dan diagnosa pasien[2]. Skema *Lazy Learning* pada metoda DeEPs sangat berguna diterapkan pada aplikasi praktis dengan data yang sering di-*update*[2]. Kemunculan pola bisa digunakan untuk atribut diskrit dan juga kontinu. Untuk atribut kontinu dibutuhkan penanganan lebih dengan mendiskritkan nilainya menggunakan metoda interseksi berbasis ketetanggaan[2].

Dalam melakukan klasifikasi DeEPs terdiri dari dua langkah utama yaitu proses pencarian pola dan proses klasifikasi. Pertama proses pencarian pola, diawali dengan mengirisakan *item-item* sebuah instan pada data *testing* dengan masing-masing *item*

instan pada data *training* untuk mereduksi jumlah *itemset* atau pola yang akan diamati. Selanjutnya memilih pola terpanjang dari masing-masing kelas pada data *training* dan direpresentasikan dalam notasi *border* agar menjadi lebih ringkas. Lalu langkah terakhir adalah mencari pola untuk klasifikasi dengan memanfaatkan algoritma JEPPRODUSER dan algoritma rekursif BORDER-DIFF. Kedua proses klasifikasi, dilakukan dengan menghitung rasio kemunculan pola pada masing-masing kelas di data *training* dan membandingkannya untuk menemukan pola yang paling dominan muncul sebagai penentu kelas dari instan data *testing* yang akan diklasifikasi.

Berdasarkan uraian di atas maka diangkatlah pemanfaatan DeEPs sebuah klasifier berbasis instan untuk mencari dan memanfaatkan kemunculan pola pada studi kasus data gejala fisik pasien demam tifoid pada tugas akhir ini. DeEPs diharapkan dapat menjadi salah satu solusi alternatif metoda klasifikasi berbasis instan yang ada saat ini.

## 1.2 Perumusan masalah

Permasalahan yang akan diangkat pada tugas akhir ini adalah bagaimana membangun Klasifier DeEPs untuk memprediksi data gejala fisik penderita demam tifoid dalam upaya menganalisis performansi DeEPs dari segi akurasi dan skalabilitas sebagai sebuah klasifier.

## 1.3 Tujuan

Berdasarkan rumusan masalah di atas maka tujuan tugas akhir ini adalah:

1. Mengimplementasikan metoda DeEPs untuk klasifikasi data gejala fisik demam tifoid dalam sebuah aplikasi perangkat lunak.
2. Menganalisis performansi aplikasi yang dibangun dari segi akurasi, waktu dan skalabilitas jumlah record.

Hipotesa : Penerapan metoda klasifikasi DeEPs pada studi kasus data gejala fisik penderita demam tifoid diharapkan akan memberikan performansi yang baik karena DeEPs sebelumnya telah terbukti dari eksperimen pada referensi terkait mampu menangani data untuk kasus klasifikasi dengan variasi dan volume data yang lebih besar.

#### 1.4 Batasan Masalah

Berdasarkan rumusan dan tujuan masalah maka batasan masalah pada tugas akhir ini adalah :

- a) Data *training* dan data *testing* adalah data pasien Rumah Sakit Bina Sehat lokal di Dayeuh Kolot Kabupaten Bandung pada tahun 2009. Berupa data pasien yang termasuk dalam kriteria diduga terkena tipus. Sesuai dengan data yang telah didapatkan pada proses pengumpulan data oleh pihak ketiga[6].
- b) Ujicoba kemampuan aplikasi yang dibangun menggunakan metoda DeEPs terhadap dataset yang bervariasi dalam jumlah *record* dan jumlah kelas (pada 2 dan 4 kelas) berdasarkan studi kasus gejala fisik penderita demam tifoid.

#### 1.5 Metodologi penyelesaian masalah

Metode yang digunakan dalam penyelesaian tugas akhir ini adalah menggunakan metode studi pustaka atau studi literatur dan analisis dengan langkah kerja sebagai berikut :

1. Studi Literatur :  
Tahapan kegiatan penambahan wawasan dengan mencari dan mempelajari topik-topik yang berhubungan dan layak dengan tugas akhir ini melalui buku referensi, jurnal dan literatur-literatur penunjang lainnya,
2. Perumusan masalah dan analisis kebutuhan:  
Tahap mengenali masalah yang dikaji hasil dari studi literatur dan menganalisis kebutuhan dari perangkat lunak yang akan dibangun pada tugas akhir ini,
3. Pengumpulan data, dilakukan untuk menunjang pembangunan dan pengaplikasian perangkat lunak pada tugas akhir ini,
4. Perancangan perangkat lunak dengan menggunakan UML dan perancangan skenario pengujian,
5. Melakukan implementasi perancangan perangkat lunak yang telah dibuat,
6. Melakukan pengujian perangkat lunak yang dibangun untuk mengetahui performansi DeEPs ,  
Parameter performansi: akurasi  

$$accuracy = \frac{t\_pos}{pos} * \frac{pos}{pos+neg} + \frac{t\_neg}{neg} * \frac{neg}{pos+neg} = \frac{t\_pos + t\_neg}{pos+neg} \dots[1.1]$$
7. Analisis dan pengambilan kesimpulan serta penyusunan laporan tugas akhir.

## 1.6 Sistematika Penulisan

Tugas akhir ini terdiri dari lima pokok pembahasan yang terangkum dalam bab masing-masing sebagai berikut :

**BAB I           PENDAHULUAN**

Pada bab ini dijelaskan latar belakang, perumusan, tujuan dan hipotesa penelitian, pembatasan, metodologi penyelesaian masalah serta sistematika penulisan.

**BAB II           DASAR TEORI**

Pada bab ini dijelaskan tentang dasar teori dari klasifikasi berbasis instan menggunakan *Emerging Pattern* atau DeEPs.

**BAB III          ANALISIS DAN PERANCANGAN SISTEM**

Pada bab ini diuraikan mengenai gambaran umum, efisiensi algoritma, batasan dan pemodelan yang digunakan untuk membangun sistem.

**BAB IV          EVALUASI HASIL**

Bab ini menguraikan hasil pengujian metoda klasifikasi DeEPs.

**BAB V           KESIMPULAN DAN SARAN**

Berisi kesimpulan dari hasil penelitian tugas akhir ini serta saran-saran untuk pengembangan lebih lanjut

## 5 KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan pengujian dan analisis maka dapat diambil beberapa kesimpulan :

1. Metoda DeEPs dapat diimplementasikan pada studi kasus klasifikasi penyakit tipus seperti yang sudah dijelaskan pada proses implementasi dan pengujian sistem.
2. Penguraian *powerset* pada BORDER-DIFF secara *brute force* membutuhkan waktu dan biaya sumber daya komputasi yang sangat besar. Sehingga perlu dilakukan efisiensi guna menekan jumlah EP yang harus diuraikan untuk pembuatan fungsi solusi lokal.
3. Dari perbandingan performansi akurasi klasifikasi untuk data 2 kelas dengan 4 kelas pada tugas akhir ini, ternyata ketepatan prediksi metoda DeEPs pada data 2 kelas lebih baik dibanding data 4 kelas. Dengan nilai performansi akurasi maksimum pada data 2 kelas adalah 96% sedangkan performansi akurasi maksimum pada data 4 kelas adalah 83.3%.
4. Hasil Pengujian pada data studi kasus klasifikasi penyakit tipus menunjukkan penambahan jumlah data *training* akan meningkatkan akurasi prediksi dari metoda DeEPs baik pada data 2 dan 4 kelas.
5. Dari perbandingan performansi waktu klasifikasi untuk data 2 dan 4 kelas pada tugas akhir ini. Waktu prediksi metoda DeEPs pada data 2 kelas lebih cepat dibanding pada 4 kelas. Dengan waktu klasifikasi saat akurasi terbaik pada data 2 kelas adalah 2.3 detik sedangkan waktu klasifikasi saat akurasi terbaik pada data 4 kelas adalah 15.3 detik.
6. Dari pengujian pengaruh peningkatan data *training* terhadap waktu klasifikasi. Ternyata pada data 2 kelas, waktu klasifikasi akan menurun dari jumlah *training* 100 sampai dengan 300 data. Dan kemudian naik pada jumlah *training* 400 data. Penurunan waktu klasifikasi pada jumlah *training* 100 sampai 300 data terjadi karena penurunan jumlah EP yang harus diuraikan untuk pembuatan fungsi solusi lokal.
7. Pengujian pengaruh peningkatan data *training* terhadap waktu klasifikasi pada data 4 kelas menunjukkan waktu klasifikasi akan naik seiring dengan pertambahan jumlah data *training*.

## 5.2 Saran

Untuk meningkatkan program dari segi *hardware* maupun *software*, penulis menyarankan agar:

1. Sebaiknya data studi kasus yang diberikan untuk tugas akhir yang serupa, menggunakan dataset yang lebih banyak lagi agar mengoptimalkan kinerja algoritma yang digunakan.
2. Sebaiknya ditambahkan penanganan terhadap atribut bertipe kontinu untuk melihat pengaruh terhadap metoda DeEPs. Dapat menggunakan metoda berbasis ketetanggan seperti yang sebutkan pada referensi utama tugas akhir ini[2].
3. Perlunya melakukan analisis lebih lanjut dari faktor yang mempengaruhi lama prediksi metoda DeEPs, agar hasil performasi metoda tidak hanya dominan dipengaruhi data.



## DAFTAR PUSTAKA

- [1] Fan, H., 2004, *Efficient Mining of interesting Emerging Patterns and Their Effective Use in classification*, The University of Melbourne, Victoria-Australia.
- [2] Li, J., Dong, G., Ramamohanarao & K., Wong, L., 2004, DeEPs: A New Instance-Based Lazy Discovery and Classification System. *Machine Learning* 54(2), pp.99-124.
- [3] Bailey, J. & Dong, G., 2007, *Contrast Data Mining: Methods dan Applications*.
- [4] Li, J., Ramamohanarao, K. & Dong, G., 2000, The space jumping Emerging Pattern an its increamental maintenance algoritm, *Proceedings of seventeenth international conference of machine learning*, pp.551-558, Stanford, CA, USA, San fransisco, morgan kaufmaan.
- [5] Li, J., Ramamohanarao, K. & Dong, G., 2001, combining the strength pattern frequency and distance for classification, *The fifth Pacific-asia conference on knowledge discovery and data mining*, pp.455-466, Hong kong.
- [6] Arifyan, A., 2010, *Analisis Implementasi dan Perbandingan decision tree J48 dengan JST Quickprop pada Studi kasus Teknik Klasifikasi dan Prediksi Penyakit Demam Tifoid(Tipus)*, ITTelkom, Bandung.
- [7] Han, J. & Kamber, M., 2006, *Data Mining: Concepts and Techniques Second Edition*, 500 Sansome Street, Suit 400, San fransisco, morgan kaufmaan.
- [8] Onder, O., 2005, *Extending DeEPs, An Instance-Based Lazy Discovery And Classification System*, Bilkent University, Ankara – Turkey.
- [9] Witten, H.I. & Frank, E., 2005, *Data Mining: Practice machine learning tools and techniques Second Edition*, 500 Sansome Street, Suit 400, San fransisco, morgan kaufmaan.
- [10] Santosa, B., 2007, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Yogyakarta, Graha Ilmu.
- [11] Hendrickx, I., & Van den Bosch, A., 2005, Hybrid algorithms with Instant-Based Classification, *Machine Learning: ECML2005*, pp. 158-169, Springer, [http://en.wikipedia.org/wiki/Eager\\_learning](http://en.wikipedia.org/wiki/Eager_learning), 13 November 2010, diakses 2 Maret 2011.

- [12] Bontempi, G., Birattari, M. & Bersini, H., 1999, *Lazy Learning Modeling & Control*, [http://en.wikipedia.org/wiki/Lazy\\_learning](http://en.wikipedia.org/wiki/Lazy_learning), 31 January 2010, diakses 2 Maret 2011.
- [13] Bremner, D., Demaine, E., Erickson, J., Iacono, J., Langerman, S., Morin, P., Toussaint, G., 2005, Output-sensitive algorithms for computing nearest-neighbor decision boundaries, *Discrete and Computational Geometry* 33(4), pp.593–604, [http://en.wikipedia.org/wiki/K-nearest\\_neighbor\\_algorithm](http://en.wikipedia.org/wiki/K-nearest_neighbor_algorithm), 12 February 2011, diakses 2 Maret 2011.
- [14] Russell, S. & Norvig, P., 2003, *Artificial Intelligence: A Modern Approach*, second edition, p.733. Prentice Hall, [http://en.wikipedia.org/wiki/Instance-based\\_learning](http://en.wikipedia.org/wiki/Instance-based_learning), 2 August 2010, diakses 2 Maret 2011.
- [15] Booch, G., Rumbaugh, J. & Jacobson, I., 1999, *The Unified Modeling Language User Guide*, Addison Wesley Longman, Inc.
- [16] Dong, G. & Li, J., 1999, Efficient mining of Emerging pattern: Discovering trends and differences, in *Proc. 1999 Intl. Conf. Knowledge Discovery and Data Mining (KDD '99)*, pp.43-52, San Diego, CA, USA.