

ANALISIS DAN IMPLEMENTASI MAX DIFF HISTOGRAM DAN ALGORITMA GEN MAX DALAM DETEKSI KECURANGAN PRESENSI KARYAWAN

Ahmad Zainur Rhofiqin¹, Toto Suharto², Kusuma Ayu Laksitowening³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Presensi karyawan merupakan salah satu masalah yang sering dihadapi oleh HRD perusahaan. Selain merupakan aspek penilaian kinerja pegawai, presensi juga merupakan suatu nilai ukur terhadap kejujuran, kedisiplinan dan profesionalitas kerja karyawan dalam suatu perusahaan.

Dewasa ini memang telah banyak mesin presensi yang canggih dan mudah digunakan. Mesin presensi tersebut antara lain mesin presensi pembaca kartu barcode, mesin presensi pembaca sidik jari dan lain sebagainya. Akan tetapi kesemua mesin presensi tersebut tidak bisa digunakan untuk mendeteksi dan menemukan adanya pola kecurangan bila ada karyawan yang menitipkan kartu presensinya untuk diisikan oleh karyawan lainnya sehingga dapat berangkat dan pulang kerja bukan di waktu yang ditetapkan.

Di Tugas Akhir ini, akan dilakukan proses analisis dan pengimplementasian metode MaxDiff Histogram dan Algoritma GenMax untuk mendeteksi pola presensi pegawai dan menemukan adanya kecurangan presensi pegawai dalam suatu perusahaan. Data yang akan diteliti adalah data presensi karyawan yang direkap oleh bagian HRD dalam suatu rentang waktu tertentu.

Hasil keluaran sistem berupa daftar karyawan yang diduga curang, association rule yang dihasilkan, persentase frekuensinya beserta waktu terjadinya kecurangan. Informasi tersebut dapat dimanfaatkan oleh pihak HRD untuk memutuskan kebijakan lebih lanjut terkait masalah presensi karyawan

Kata Kunci : Presensi, Max Diff Histogram, Association rule, GenMax

Abstract

Presence of employees is one of the problems often faced by HRD in many company. Beside to measure the performance of employee, presence is also a measure of the value of honesty, discipline and professionalism of employees in a company.

Today, there are some sophisticated and easy to use machines presence, such as barcode card readers, fingerprints readers and etc. But all of this machine can not be used to detect and find a pattern of cheating when there are employees who left their presence card to be filled by the other employees, so that they can go to work not in the set time.

This Final Project will perform an analysis and implementation of MaxDiff Histogram and GenMax Algorithm to detect pattern and find the cheated employee in the company. The data to be examined is the presence employee data recapitulated by the HRD department in some period.

The output of this system is a list of employees who are suspected of cheating, the association rule and the percentage frequency of fraud. This information can be used by the HRD division to decide on further policy issues related to employee.

Keywords : Presence, Max Diff Histogram, Association rule, GenMax

1. Pendahuluan

1.1 Latar Belakang

Dalam sebuah perusahaan, masalah presensi dan kehadiran karyawan merupakan suatu persoalan yang agak rumit. Kedua hal di atas menjadi sangat penting karena terkait dengan aspek kinerja, tingkat kejujuran dan tingkat kedisiplinan seorang karyawan dalam suatu perusahaan.

Meskipun dewasa ini sudah ada beberapa alat dan *tools* untuk mencatat kehadiran karyawan, seperti Mesin Presensi Elektronik (*Check-lock*), mesin presensi sidik jari dan kartu barcode, akan tetapi kesemua alat tersebut mempunyai satu kekurangan, yaitu tidak bisa mendeteksi dan menemukan adanya kecurangan bila ada karyawan yang menitipkan kartu presensinya untuk diisikan oleh karyawan lainnya sehingga dapat berangkat dan pulang kerja bukan di waktu yang ditetapkan. Dan hal itu menjadi suatu persoalan tersendiri bagi divisi HRD dalam suatu perusahaan [10].

Di penelitian ini, untuk mendeteksi adanya kecurangan dalam kegiatan presensi, data rekap presensi selama rentang waktu tertentu, misalnya dalam rentang waktu satu bulan, akan direpresentasikan ke dalam bentuk Max Diff Histogram sehingga didapatkan pola persebaran dan kelompok (*bucket*) datanya. Histogram tersebut kemudian diproses dengan algoritma GenMax untuk menggali Maksimum Frequent Itemset. Hasil Maksimum Frequent Itemset di atas akan diproses lebih lanjut untuk menentukan pola *asosiation rule* yang ada, sehingga akan didapatkan daftar kandidat karyawan yang kemungkinan besar melakukan kecurangan presensi. Disini diasumsikan bahwa kecurangan presensi terjadi bila ada dua orang karyawan atau lebih yang melakukan presensi kehadiran dalam jangka waktu yang sangat berdekatan (dalam satuan detik) dalam suatu kurun waktu tertentu. Sehingga besar kemungkinan bahwa hanya seorang karyawan yang melakukan proses presensi sedangkan karyawan yang lain hanya menitipkan kepada salah seorang karyawan tersebut.

Dalam penelitian ini, digunakan Metode Max Diff Histogram karena model histogram ini dirasa paling tepat untuk merepresentasikan pola persebaran data yang akan diteliti. Selain itu, Max Diff Histogram juga memberi batasan yang jelas antar satu bucket (kelompok data) dengan bucket yang lainnya, yaitu sebesar $(\beta - 1)$ [6]. Di sini, nilai β ditentukan dalam satuan detik yang menunjukkan waktu minimum selisih 2 orang untuk melakukan presensi secara normal. Sehingga dengan adanya batasan di atas akan meminimalisasi jumlah *bucket* data dan akan mempersempit ruang pencarian solusi.[10]

Algoritma yang digunakan untuk memilih *Maximum Frequent Itemset* adalah Algoritma GenMax. Alasan digunakannya algoritma ini adalah karena Algoritma GenMax menggunakan metode pencarian secara *backtracking* atau runut balik, sehingga efisien untuk mengenumerasi semua kandidat maksimal pattern yang ada. Selain itu, Algoritma GenMax menggunakan teknik *progressive focusing*, yaitu teknik untuk menghilangkan *non-maximal frequent itemset*, yang membuat ruang pencarian menjadi lebih sempit dan teknik *diffset propagation* yang bertujuan mempercepat komputasi frekuensi [4].

Diharapkan dengan penggunaan gabungan metode di atas hasil keluaran system diharapkan mempunyai akurasi yang tinggi, dan memerlukan waktu yang singkat. Hasil analisis yang didapat dari proses ini adalah berupa daftar beberapa

karyawan yang diduga melakukan kecurangan presensi beserta prosentasenya yang akan disampaikan kepada divisi HRD perusahaan yang bersangkutan. Informasi diatas dapat digunakan oleh pihak HRD untuk memutuskan langkah-langkah dan kebijakan lanjut terkait penanganan beberapa karyawan yang diduga melakukan kecurangan presensi tersebut.[10]

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan diatas, maka masalah yang akan dirumuskan dalam Penelitian Tugas Akhir ini adalah sebagai berikut :

1. Bagaimana kinerja/performa Max Diff Histogram dan Algoritma GenMax dalam proses pendeteksian kecurangan presensi pada data presensi karyawan suatu perusahaan?
2. Bagaimana pengaruh nilai β (nilai pembeda jarak antar bucket) pada pembentukan Histogram Max Diff, parameter `min_sup` dan `min_conf` terhadap akurasi hasil akhir sistem ini ?

1.3 Batasan Masalah

Dalam implementasi tugas akhir ini dibatasi oleh beberapa hal, sebagai berikut :

1. Data yang akan diuji dan dianalisis adalah data presensi karyawan yang dicatat dengan mesin presensi elektronik (baik itu mesin check-lock, barcode, kartu ataupun mesin lainnya)
2. Hanya data presensi yang lengkap dan valid saja yang akan diolah dan diteliti
3. Data yang akan digunakan adalah data karyawan hasil rekap divisi HRD PT Bina Citra Kharisma Lestari Bandung dan PT Dhanar Mas Concern dalam kurun waktu tertentu (misalnya 1-3 bulan)
4. Hasil outputan adalah daftar karyawan yang diduga melakukan kecurangan presensi beserta prosentase frekuensinya
5. Data yang akan diteliti adalah data presensi karyawan, yang dicatat dengan menggunakan 1 mesin presensi dari 1 bagian / divisi.

1.4 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah :

1. Mengimplementasikan Metode Max Diff Histogram dan algoritma Gen Max untuk mendeteksi kecurangan data presensi kehadiran karyawan pada suatu perusahaan
2. Mengetahui pengaruh nilai β (nilai pembeda jarak antar bucket), `min_support` dan `min_confidence` terhadap akurasi hasil akhir sistem yang dibangun.

1.5 Metode Penyelesaian Masalah

Metodologi / langkah-langkah penyelesaian masalah yang akan digunakan adalah sebagai berikut :

1. Studi literatur

Merupakan tahapan dalam mempelajari konsep dan teori pendukung untuk memecahkan permasalahan. Dalam pengerjaan tugas akhir ini, studi literatur meliputi pembelajaran teknik,-teknik preprosesing

data, konsep histogram, terutama Max Diff Histogram, algoritma MaxGen serta informasi lainnya yang menunjang pembuatan tugas akhir ini.

2. Pengumpulan data

Pada tahap ini, akan dilakukan proses pencarian data presensi karyawan suatu perusahaan yang nantinya akan dianalisis dan diuji sesuai permasalahannya

3. Pemodelan dan pembuatan sistem

Pemodelan dan pembuatan sistem yang akan dibangun meliputi hal-hal sebagai berikut :

- a) Pemodelan preprosesing data sumber
- b) Pemodelan proses pembuatan histogram
- c) Pemodelan proses pemilihan Maksimum Frequent Itemset dan proses pencarian Association Rule
- d) Pemodelan dan pembuatan antar muka / interface sistem
- e) Pemodelan pengujian dan proses analisa hasil

4. Testing dan analisa hasil

Melakukan pengujian aplikasi dan menganalisis hasil keluaran aplikasi. Di tahap testing ini, nilai parameter beta, min_support dan min_confidence terhadap hasil akurasi sistem.

5. Pembuatan laporan Tugas Akhir

Melakukan proses penyusunan laporan hasil penelitian yang telah dilakukan serta memberikan kesimpulan dari hasil penelitian yang telah dilakukan

5. Penutup

5.1 Kesimpulan

Berdasarkan analisis yang diperoleh dari hasil pengujian dalam penelitian Tugas Akhir ini, dapat ditarik kesimpulan sebagai berikut :

2. Gabungan antara MaxDiff Histogram dan Algoritma GenMax dapat diterapkan dan diimplementasikan untuk proses deteksi kecurangan presensi karyawan yang dicatat dengan mesin presensi elektronik dengan menghasilkan akurasi yang tinggi.
3. Parameter beta (jarak pembeda antar bucket) merupakan parameter yang paling utama dalam sistem ini, karena parameter ini merupakan pembeda utama antara karyawan yang diduga melakukan kecurangan presensi dan karyawan yang berangkat dan pulang kerja bersama-sama.
4. Parameter min_support berpengaruh terhadap jumlah item yang akan diproses dengan Algoritma GenMax untuk pencarian MFI. Semakin besar nilai support semakin kecil jumlah item yang akan diproses oleh algoritma GenMax
5. Parameter min_confidence akan berpengaruh terhadap jumlah kandidat rule yang terpilih pada proses pemilihan rule pada tahap *association rule*.

5.2 Saran

Berdasarkan hasil analisis dan kesimpulan, saran untuk perbaikan pada penelitian deteksi kecurangan presensi karyawan adalah sebagai berikut :

1. Sebaiknya data masukan sistem dirancang sedemikian rupa sehingga sesuai dengan format data keluaran mesin presensi, sehingga proses *data structuring* bisa dihilangkan.
2. Sebaiknya data verifikasi yang didapat dari bagian HRD perusahaan harus sevalid mungkin sehingga dapat digunakan untuk mengukur performansi sistem yang telah dibangun.

Telkom
University

Daftar Pustaka

- [1] Buccafurri, Francesco, Gianluca Lax, Domenico Sacca, Luigi Pontieri and Domenico Rosaci, "Enhancing Histograms by Tree-Like Bucket Indices", Proceeding of the International Conference on Data Engineering 2002, IEEE Computer Society, 2002
- [2] Burdick, D., Calimlim, M., and Gehrke, J. 2001. MAFIA: A maximal frequent itemset algorithm for transactional databases. In IEEE Intl. Conf. on Data Engineering, pp. 443–452.
- [3] Gouda, Karam, Mohammed J Zaki, "Fast Vertical Mining Using Diffsets", Portal ACM : 2003
- [4] Gouda, Karam, Mohammed J Zaki, " GenMax: An Efficient Algorithm for Mining Maximal Frequent Itemsets", Data Mining and Knowledge Discovery, Springer Science 2005
- [5] H. V. Jagadish Nick Koudas S. MuthViswanat, uriskhant, h Poosala, Ken Sevcik, Torsten Sue, "Optimal Histograms with Quality Guarantees", Bell Labs Technical Report, May 1998
- [6] Han, Jiawei, Micheline Kamber, "Data Mining: Concepts and Techniques", Second Edition, Morgan Kaufmann, 2006
- [7] Hernandez, Raudel, Airel Perez Suarez, "An Algorithm for Mining Frequent Itemsets", 2008 5th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE 2008)
- [8] Huang Kuo-Yu, "The Concept of Maximal Frequent Itemsets", NCU CSIE Database Laboratory, May 2002
- [9] Poosala, Viswanath, Yannis E. Ioannidis Peter, J. Haas Eugene, J. Shekita, "Improved Histograms for Selectivity Estimation of Range Predicates", ACM SIGMOD Conference 1996, Montreal Canada
- [10] Satia, Budhi, Gregorius, Soedjianto Felicia, "Aplikasi Data Mining Market Basket Analysis pada Tabel Data Absensi Elektronik untuk Mendeteksi Kecurangan Absensi (Checkclock) Karyawan di Perusahaan", Journal UK Petra Surabaya Vol 8 / No 2, 2007-Research Center
- [11] <http://en.wikipedia.org/wiki/Histogram> diakses pada tanggal 16 april 2011

Telkom
University