**MANAJEMEN TEKNOLOGI INFORMASI**

ANALISIS PENERAPAN KNOWLEDGE DISCOVERY IN DATABASES

DALAM MENENTUKAN KINERJA PEGAWAI

****

**MIRA AFRIANA UTAMI**

**09031281621044**

**Dosen Pengampu : Deris Stiawan, Ph.D**

**JURUSAN SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2018**

**Analisis Penerapan KDD Dalam Menentukan Kinerja Pegawai**

Sumber daya manusia merupakan salah satu faktor strategis perusahaan yang memberikan nilai tambah sebagai tolak ukur keberhasilan bisnis suatu perusahaan. Kemampuan SDM merupakan competitive advantage dari perusahaan. Untuk dapat menjadi daya saing utama, pembentuk daya saing perlu melakukan upaya peningkatan kinerja sumber daya manusia. Selain itu, rekrutmen pegawai baru adalah salah satu proses bisnis paling penting yang mempengaruhi kualitas SDM dalam setiap perusahaan. Sangat penting bagi perusahaan untuk memastikan perekrutan bakat yang tepat guna mempertahankan keunggulan kompetitif atas yang lain di pasar.

**Apa itu KDD (Knowledge Discovery in Databases) ?**

 KDD (Knowledge Discovery in Databases) adalah proses terorganisir untuk mengidentifikasi pola yang valid, baru, berguna, dan dapat dimengerti dari sebuah dataset yang besar dan kompleks. Salah satu tahapan penting dalam proses KDD adalah Data Mining.

Tahapan Knowledge Discovery in Databases (KDD) yang terdiri dari :



*Tahapan Knowledge Discovery in Database*

1. *Data Selection,* pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.
2. *Preprocessing/Cleaning,* sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahn cetak*.*
3. *Transformation,* coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada sejenis atau pola informasi yang akan dicari dalam basis data.
4. *Data Mining,* data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
5. *Evaluation,* pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

**Analisis Jurnal/Paper :**

**1. (Dhika, 2015)**

 Penelitian yang dilakukan penulis, pengambilan keputusan kenaikan jabatan menggunakan metode KDD (knowledge Discovery in Databases) dan Metode Algoritma C4.5 dimana dalam pengolahan data akan dilakukan menggunakan software pendukung yaitu Rapidminer versi 5.1. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Sedang pohon keputusan dapat diartikan suatu cara untuk memprediksi atau mengklarifikasi yang sangat kuat. Pohon keputusan dapat membagi kumpulan data yang besar menjadi himpunan.



Metode dan tahapan penelitian menggunakan metode *Knowledge Discovery in Database* (KDD) yaitu meliputi;

1. *Domain Understanding & KDD Goal,* yakni dilakukan pemahaman terlebih dahulu hasil yang akan dicapai, bahasan dan bidang yang akan dikerjakan.
2. *Selection & Addition*, merupakan tahapan memilih dan menambahkan data yang diperoleh dari objek penelitian.
3. *Preprocecing*, merupakan proses data awal yang sudah diperoleh, melalui pemilihan data yang lengkap, tepat dan sesuai.
4. *Transformation*, merupakan perubahan *type text* ke dalam type yang susuai dengan kebutuhan algoritma terutama penggunaan algoritma C4.5
5. *Data mining*, merupakan proses penerjemahan data dimana dilakukan penerapan algoritma yang digunakan dalam pengolahan data. Pada bagian ini disesuaikan langkah pada algoritma C4.5
6. *Evaluation dan Interpretation*, merupakan tahapan evaluasi dan penterjemahan data yang diperoleh dari algoritma C4.5
7. *Discovered Knowledge*, merupakan pemanfaatan hasil data yang telah menjadi *rule* baru yang akan diimplementasikan, implementasikan dalam visual.

Hasil dari pengujian model yang telah dilakukan dengan algoritma C4.5 yaitu dilakukan pengujian tingkat akurasi dengan menggunakan *confussion matrix* dan kurva ROC/AUC (*Area Under Cover*) diperoleh melalui Rapidminer dengan menginput data dari PT. ABC.

Dari Hasil penelitian tersebut, dapat disimpulkan bahwa : Akurasi data dapat diklasifikasikan sebagai *excellent classification* nilai yang di peroleh 0.930 sehingga rule yang dihasilkan dapat diterapkan dan digunakan dalam perusahaan sebagai acuan dalam menentukan kenaikan jabatan karyawan.

**2. (Sharma, Vishwakarma, & Lakhtaria, 2013)**

Penilaian kinerja sumber daya manusia adalah tugas yang sensitif. Untuk menghindari keberpihakan, alat yang efisien untuk menangani berbagai data dan membantu manager untuk membuat keputusan serta rencana adalah hal yang luar biasa membantu. Dalam data mining, data historis seperti atribut-atribut yang mempengaruhi kinerja dapat dimanfaatkan sebagai pengalaman pembelajaran. Ini dapat digunakan untuk memprediksi keadaan masa depan dan sumber pengetahuan yang kaya dan mendukung keputusan. Secara umum, untuk melakukan proses discovering information perlu menggambarkan beberapa prediksi-prediksi ilmiah. Kerangka kerja dibutuhkan sebagai sebuah panduan untuk menelusuri detail dan prosedur.



Pada penelitian ini akan digabungkan metode CRISP-DM dengan Proses Knowledge Dicovery in Databases dan menggunakan tools WEKA. Penelitian ini menggunakan aturan berbasis teknik klasifikasi menggunakan penutup sekuensial algorithm (SCA) dan metode hold-out dari partisi set data. Untuk kualitas aturan, FOIL dipilih karena mempertimbangkan cakupan dan akurasi. Di WEKA, Algoritma kloning Ripper yang disebut Jrip dirancang untuk mengeksekusi klasifikasi set data saat mensimulasikan proses algoritma penutupan sekuensial. kemampuannya untuk langsung mengklasifikasikan set data tanpa harus mendasarkan aturan pada pohon keputusan, memberikan bukti kesederhanaannya. Dalam Penelitian ini, prediksi dibuat pada kinerja karyawan yang baru diangkat berdasarkan pelatihan mereka saat ini.

**3. (Mukhibah & Kurniawati, 2017)**

 Pada penelitian ini menggunakan *Cross Industry Standard Process for Data Mining Methodology* (CRISP-DM).

Terdapat 6 tahapan :

1. *Business Understanding/* Memahami tujuan bisnis, yaitu membantu PT. Emerio Indonesia dalam mene-mukan kriteria rekruitmen *Software Engineer* yang berkualitas.
2. *Data Understanding/* Pemahaman data, yaitu mengacu pada dokumen data personal dan dokumen penilaian *performance Software Engineer* di PT. Emerio Indonesia.
3. *Data Preparation/* Persiapan data, yaitu membuat format data dalam bentuk tabular dengan mengguna-kan excel dari dokumen *applicationform* dan *performance appraisal* untuk kemudian dikonversi ke format csv.
4. *Modelling/*Pemodelan yaitu mem-buat model klasifikasi berdasarkan data csv *Software Engineer* dengan menggunakan *tools* WEKA dengan metode *decision tree*.
5. *Evaluation/*Evaluasi, yaitu menge-valuasi hasil model klasifikasi apakah sudah sesuai dengan tujuan penelitian.
6. *Deployment*/Penyebaran,pem-buatan laporan tentang pengetahuan yang didapatkan dari hasil implementasi *data mining* terhadap prediksi *performanceSoftware Engineer*.

Kriteria penilaian yang digunakan untuk menilai *performance Software Engineer* dibagi menjadi dua kategori yaitu kemampuan teknis dan kemampuan non-teknis. Pada tahap Pemahaman ini, implementasi *data mining* di PT. Emerio Indonesia ditujukan untuk mendapat-kan model *decision tree*yang meng-gambarkan pola hubungan antara kemampuan teknis dan non-teknis dengan *performance* Software Engineer di divisi *software service.* Model *decision tree* ini dapat dijadikan sebagai acuan dalam melakukan proses rekruitmen *Software Engineer* yang tepat. Kemudian pada tahap pemahaman data, data dikumpulkan dengan cara mencari informasi tentang data per-sonal *Software Engineer* dan hasil *performance appraisal* melalui depar-temen *Human Resource* (HR) PT. Emerio Indonesia. Pada tahap pengolahan data, dilakukan proses untuk membentuk ***dataset* “*performance*”** dengan menggunakan Microsoft Excel 2010. Kemudian dilakukan transformasi data, langkah terakhir dari proses pengolahan data adalah mengubah *dataset* dari *file* excel menjadi *format* CSV agar dapat dikenali sebagai sumber data pada WEKA. Pengolahan data pada WEKA merupakan tahap untuk memilih sumber data pemodelan dari hasil persiapan data.

Selanjutnya adalah Tahap Pemodelan pada WEKA merupakan tahap untuk memodelkan prediksi *performanceSoftware Engi-neer* dengan menggunakan algoritma *decision tree* C4.5. Proses pemodelan data dengan klasifikasi *decision tree*. Berdasarkan hasil pemodelan dengan WEKA diperoleh jumlah *leave* dari model klasifikasi *decision tree* untuk *performance* Software Engineer adalah tujuh, ukuran dari *tree* adalah 13 dan sebagai *root* atas *decision tree* tersebut adalah atribut desain dengan nilai 2.85. R*ule* yang didapatkan dari hasil pemodelan. Namun, pada tahap terakhir yaitu Penyebaran tidak dapat dilakukan karena kebijakan internal PT. EMERIO INDONESIA.

 Dari hasil penelitian tersebut dapat disimpulkan bahwa berdasarkan data yang diproses untuk pemodelan prediksi *performance Software Engineer* di PT. Emerio Indonesia, dihasilkan 98 *instance* dengan klasifikasi benar dan 3 *instance* dengan klasifikasi salah. Sehingga tingkat akurasi dari pemodelan tersebut adalah 97.0297% dan tingkat kesalahan sebesar 2.9703%. Dengan tingkat kesalahan yang relatif kecil ini, dapat disimpulkan bahwa model klasifikasi *decision tree* pada imple-mentasi data mining dalam prediksi *performance Software Engineer* di PT. Emerio Indonesia dapat digunakan sebagai acuan dalam membuat ke-bijakan proses perekrutan *Software Engineer*yang lebih baik. Perekrutan *Software Engineer* ini melalui kriteria-kriteria *Software Engineer* yang dihasilkan oleh pemodelan, yaitu *very competent, competent* dan *development needed*.

**4. (Firdaus, 2016)**

 Data mining merupakan tahapan untuk menemukan pola atau informasi dalam sekumpulan data dengan menggunakan teknik dan algoritma tertentu. Pemilihan teknik dan algoritma yang tepat sangat bergantung pada proses KDD secara keseluruhan. Pada penelitian ini penerapan data mining menggunakan metode Decision Tree (J4S) dan Naive Bayes untuk menentukan klarifikasi kinerja karyawan.

Data yang akan di-mining diproses melalui tahapan Knowledge Discovery in Databases (KDD) dengan menggunakan SQL Server. Tahap pertama yaitu Data Selection. Data yang digunakan dalam penelitian ini berasal dari data eksperimen penulis yaitu data pegawai, terdiri dari beberapa tabel antara lain tabel karyawan dan tabel peniaian. Tabel karyawan berisi tentang informasi data pegawai dan tabel penilaian berisi tentang informasi data kehadiran, data prestasi kaywan dan data tunjangan karyawan . Jumlah *dataset* pada data penilaian sebanyak record dalam 12 bulan. Pada tahapan kedua yaitu preprocessing ini akan dilakukan proses integrasi data untuk menghubungkan data karyawan dan data penilaian, selanjutnya dilakukan proses data cleaning untuk menghasilkan dataset yang bersih sehingga dapat digunakan dalam tahap berikutnya yaitu mining dengan tujuan memperoleh pola mengenai kriteria pegawai untuk menentukan klasifikasi pegawai. Pada tahapan ketiga transformation data merupakan tahap merubah data ke dalam bentuk yang sesuai untuk di-mining. Kemudian di tambahkan satu attribut untuk mempermudah proses tranformation. Dari penelitian tersebut dapat disimpulkan bahwa berdasarkan data pegawai yang diuji menggunkan SIK yang dijadikan data training, metode Naive Bayes berhasil mengklasifikasikan 49 data dari 50 data yang diuji. Sehingga dengan demikian metode Naive Bayes ini berhasil memprediksi kinerja karyawan dengan persentase keakuratan sebesar 98 %.

**5. (Al-Radaideh & Nagi, 2012)**

 Penelitian ini akan membangun model klasifikasi yang menguji atribut tertentu yang mungkin mempengaruhi kinerja pekerjaan. Untuk mencapai hal ini, Metode (CRISP-DM, 2007) digunakan untuk membangun klasifikasi model. Metode ini terdiri dari lima langkah yang meliputi: Pemahaman Bisnis, Pemahaman data, Persiapan data, Pemodelan, Evaluasi dan Penyebaran. Setelah data siap dan model klasifikasi dibangun. Langkah selanjutnya adalah mengukur rasio yang digunakan untuk menunjukkan berat keefektifan setiap atribut pada kelas yang diuji, dan sesuai urutan node pohon ditentukan. Setelah semua tahapan dilakukan. Ditemukan bahwa yang paling berpengaruh pada kinerja karyawan adalah jabatan. Namun, yang mengejutkan, dalam eksperimen E2, sebuah tren aneh muncul mengenai jumlah anak-anak, yang menunjukkan bahwa jumlah anak yang lebih tinggi mengarah ke kinerja yang lebih tinggi. Ini bisa menjadi sebuah kebetulan, karena kumpulan data E2 tidak cukup besar untuk mengkonfirmasi aturan ini. Di sisi lain jenis kelamin, tidak berlaku sama sekali pada percobaan E1 dan E2, karena proporsi perempuan di kedua dataset tidak signifikan. Namun dalam percobaan E3, itu menunjukkan kinerja yang lebih tinggi untuk karyawan pria daripada wanita. Beberapa faktor profesional juga tampaknya mempengaruhi kinerja. Gaji dalah salah satu faktor yang paling positif terhadap kinerja, efek ini telah ditunjukkan dalam percobaan E1 dan E3, sedangkan di E2 itu tidak signifikan. Jumlah sebelumnya perusahaan di E1, menunjukkan keduanya positif dan negative hubungan dengan kinerja. Ini mungkin karena karyawan yang baru bekerja tidak memiliki pengalaman bekerja dari perusahaan lain, mereka melakukan yang terbaik untuk mendapatkan posisi yang lebih baik. Akhirnya, kepuasan kerja dan lingkungan kerja yang nyaman memiliki efek pada kinerja. Untuk E3 mereka tidak dimasukkan sebagai faktor efektif; sedangkan di E1 dan E2 mereka dianggap sebagai faktor efektif berbobot rendah. Ini bisa diartikan bahwa perusahaan di E3memiliki kondisi yang lebih memuaskan daripada perusahaan di E2.Sebagai pernyataan akhir pada keakuratan klasifikasi model dibangun untuk tiga percobaan, dapat diketahui bahwa untuk algoritma yang berbeda digunakan, akurasi klasifikasi adalah lebih banyak lagi dalam eksperimen E2 dan E3 daripada di E1. Ini mungkin karena berbagai perusahaan karyawan, termasuk dalam E1, yang menciptakan berbagai faktor yang mempengaruhi kelas di percobaan. Sementara di E2 dan E3, kendati dataset mereka kecil, tetapi karyawan yang diteliti memiliki pekerjaan yang sama kondisi, lingkungan kerja, manajemen dan rekan kerja yang membuat penelitian lebih fokus pada atribut terukur di tangan..

**DAFTAR PUSTAKA**

Al-Radaideh, Q. A., & Nagi, E. A. (2012). Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance. *International Journal of Advanced Computer Science and Applications, 3*, 144.

Dhika, H. (2015, November). KAJIAN PERANCANGAN ATURAN KENAIKAN JABATAN PADA PT. ABC. *Jurnal Simetris, 6*, 219.

Firdaus, R. D. (2016). Penerapan Data Mining Untuk Menentukan Klasifikasi Kinerja Pegawai.

Mukhibah, D., & Kurniawati, A. (2017, Desember). IMPLEMENTASI DATA MINING DALAM PREDIKSI PERFORMANCE SOFTWARE ENGINEER PADA PT. EMERIO. *Informatika dan Komputer, 2*, 37.

Sharma, A. k., Vishwakarma, S., & Lakhtaria, D. K. (2013, Maret). Data Mining Based Predictions for Employees Skill Enhancement. *International Journal of Advanced Research in Computer Science, 4*(Special Issue).