

PERBANDINGAN TEKNIK KLASIFIKASI UNTUK PREDIKSI STATUS KONTRAK KERJA KARYAWAN

Fajar Priyono^{1*}, Richardus Eko Indrajit²

^{*12}Fakultas Ilmu Komputer, STMIK Nusa Mandiri, Jakarta

Jl. Damai No. 8, Warung Jati Barat (Margasatwa), Pasar Minggu, Jakarta Selatan, Indonesia 12550

*E-mail: fha.jr20@gmail.com

ABSTRAK

Teknik prediksi status kerja karyawan kontrak disebut perusahaan belum menggunakan metode yang ilmiah, hal ini menyebabkan bagian SDM tidak bisa memperkirakan faktor apakah yang paling berpengaruh terhadap keputusan status kerja karyawan kontrak akan diperpanjang ataupun diputus kontrak kerjanya dikarenakan setiap karyawan kontrak memiliki kelebihan dan kekurangannya masing-masing. Prediksi yang digunakan untuk menentukan status kerja karyawan kontrak diharapkan dapat menghasilkan suatu pengetahuan yang dapat digunakan bagian SDM dalam memutuskan status kerja karyawan kontrak di perusahaan. Penelitian ini membahas teknik klasifikasi utama yang meliputi Decision Tree, KNN, Naïve Bayes, Deep Learning dan Rule Induction. Percobaan ini dilakukan menggunakan data dari sebuah perusahaan yang berisi nama-nama karyawan kontrak dengan berbagai atributnya. Area di bawah kurva (AUC) digunakan sebagai indikator akurasi dalam rangka untuk mengevaluasi kinerja pengklasifikasi. Hasil menunjukkan bahwa metode Deep Learning menunjukkan hasil yang paling memuaskan dibandingkan metode yang lainnya.

Kata kunci: Status kerja karyawan kontrak, Klasifikasi, AUC

ABSTRACT

The prediction technique of the employment status of contract employees in a company has not used scientific method, this causes the HR department can not estimate whether the factors that have the most influence on employment status decisions of contract employees will be extended or terminated their work contracts because each contract employee has their own advantages and disadvantages. The prediction used to determine the working status of the contract employee is expected to produce a knowledge that can be used HR part in deciding the working status of contract employees in the company. This study discusses the main classification techniques that include Decision Tree, KNN, Naïve Bayes, Deep Learning and Rule Induction. This experiment is conducted using data from a company that contains the names of contract employees with various attributes. The area under the curve (AUC) is used as an indicator of accuracy in order to evaluate the performance of the classifier. The results show that Deep Learning method shows the most satisfactory result compared to other methods.

Keywords : Working status of contract employees, Classification, AUC

PENDAHULUAN

Status kerja karyawan kontrak di sebuah perusahaan merupakan hal yang penting dan perlu diperhatikan demi tercapainya tujuan perusahaan. Setiap semester pasti ada karyawan yang diputus kontrak ataupun diperpanjang kontrak kerjanya karena siklus tersebut sudah menjadi sebuah peraturan dalam sebuah perusahaan. Dalam hal untuk menentukan apakah seorang karyawan itu layak untuk diperpanjang kontrak kerjanya atau diputus kontraknya sering terjadi kesulitan dalam

menentukan hal tersebut, disamping banyaknya kriteria penilaian yang diperlukan untuk memutuskan, faktor lain seperti sikap kerja dan attitude karyawan tersebut juga perlu diperhatikan.

Penilaian hasil kerja karyawan kontrak selalu dilakukan disetiap semester, dan setiap semester pasti selalu ada karyawan kontrak yang diperpanjang ataupun diputus kontraknya. Bagian SDM hanya menerima hasil penilaian dan keputusan status kerja karyawan kontrak tersebut dari masing-masing kepala seksi per

unit stuan kerja dan bagian SDM belum mempunyai metode prediksi untuk menentukan faktor apakah yang paling berpengaruh dalam pengambilan keputusan status kerja karyawan kontrak.

Perlu suatu prediksi apakah karyawan kontrak tersebut perlu diperpanjang atau diputus status kontraknya. Untuk mengatasi hal tersebut maka digunakan teknik prediksi atau peramalan adalah teknik data mining yang digunakan dalam penelitian ini, dengan menggunakan algoritma decision tree diharapkan dapat menghasilkan suatu pengetahuan yang dapat digunakan dalam pengambilan keputusan terhadap status kerja karyawan kontrak.

Upaya bab ini untuk membuat sebuah perbandingan yang jauh lebih komprehensif dari teknik klasifikasi. Lebih khusus, teknik klasifikasi yang digunakan adalah membandingkan lima jenis model klasifikasi yang mencakup model Decision Tree (c4.5), k-NN, Naïve Bayes, Rule Induction dan Deep Learning.

Penelitian ini dilakukan di perusahaan PT. Astra Honda Motor Jakarta. Data yang digunakan adalah data penilaian karyawan kontrak yang berisi nama-nama karyawan kontrak, hasil penilaian, status kontrak dan lain sebagainya.

METODE

Business Understanding

Pada tahapan business understanding, dilakukan pemahaman terhadap objek penelitian. Pemahaman mengenai objek penelitian dilakukan dengan penggalian informasi di perusahaan PT. Astra Honda Motor, wawancara dengan pihak yang terkait dan observasi terhadap proses yang berjalan di perusahaan PT. Astra Honda Motor.

1. *Data Understanding*

Data yang digunakan dalam penelitian ini didapat dari perusahaan PT. Astra Honda Motor berupa data karyawan kontrak yang berisi 199 nama-nama karyawan kontrak beserta atribut-atribut lain yang ada didalamnya.

Atribut yang digunakan adalah:

Nama : Berisi nama-nama karyawan kontrak

Jenis Kelamin : Laki-Laki atau Perempuan

Status Karyawan : Bekerja saja atau Bekerja dan Kuliah

Umur : Umur karyawan berkisar antara 20-25

Status Pernikahan : Single atau Menikah

Nilai Mutu : Nilai penilaian kerja 1.00-5.00

Grade : Berisi Grade Nilai Mutu (IS, BS, B, C dan K)

Keputusan Kontrak : Perpanjang Kontrak atau Putus Kontrak

2. *Data Preparation*

Pada tahap data preparation merupakan tahap dengan proses penyiapan data yang bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan dalam penelitian. Tahapan yang dikerjakan adalah melakukan pengabaian atribut pada data mentah yang dianggap tidak relevan dengan hasil pengujian dan perubahan terhadap nilai data bahkan tipe data pada atribut dataset dengan tujuan untuk mempermudah pemahaman terhadap isi record dengan memperhatikan konsistensi data, missing value, dan redundancy pada data.

Data yang digunakan sudah bersih karena Missing Value=0

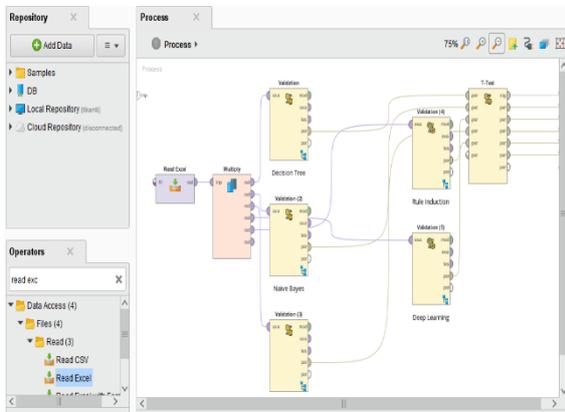
Tidak ada inkonsistensi data karena nilai average nya berada diantara nilai min dan max.

Gambar 3.1 Dataset Karyawan Kontrak
Modeling

Field	Type	Missing	Statistics	Filter (if applicable)	Search for attribute
Keputusan Kontrak	Boolean	0	Putus Kontrak (74) Perpanjang Kontrak (125) Perpanjang		
Nama	Polynomial	0	Zahab Marul (1) Aaron Benda (1) Aaron Bk		
Jenis Kelamin	Boolean	0	Perempuan (88) Laki-Laki (111) Laki-Laki		
Status Karyawan	Boolean	0	Bekerja Saja (91) Bekerja dan Kuliah (108) Bekerja & C		
Umur	Integer	0	Min: 20 Max: 25 Average: 22.759		
Status Pernikahan	Boolean	0	MENIKAH (47) SINGLE (152) SINGLE		
Nilai Mutu	Numeric	0	Min: 2.625 Max: 4.750 Average: 3.556		

Tahap ini juga dapat disebut tahap learning karena pada tahap ini data training diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Pada penelitian ini, pembuatan model menggunakan algoritma decision tree, yang merupakan salah satu metode klasifikasi. Kemudian aturan diuji lagi menggunakan data testing.

Data yang digunakan dalam penelitian ini didapat dari perusahaan PT. Astra Honda Motor berupa data karyawan kontrak yang berisi 199 nama-nama karyawan kontrak beserta atribut-atribut lain yang ada didalamnya.



Gambar 3.2 Model data

Evaluation

Pada tahap ini dilakukan pengujian terhadap model yang terbentuk untuk mendapatkan informasi keakuratan model yang dibentuk dengan validasi menggunakan metode *confusion matrix* dan kurva OC.

Deployment

Setelah pembentukan model dan dilakukan analisa dan pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini diterapkan model yang paling akurat dengan memakai data baru diluar data *training* dan data *testing*.

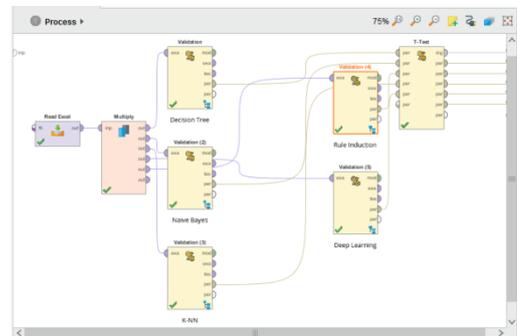
HASIL DAN PEMBAHASAN

Dataset

Pada tahap ini dilakukan eksperimen dan teknik pengujian yang digunakan untuk mebangukur tingkat akurasi algoritma berdasarkan data set status kontrak kerja karyawan yang digunakan yaitu sebanyak 199 nama karyawan beserta atribut didalamnya dari beberapa department.

Modeling

Pengolahan data menggunakan aplikasi Rapid Miner dengan menggunakan metode Cross Validation dan menggunakan 5 algoritma yaitu Decision Tree, KNN, Naïve Bayes, Deep Learning dan Rule Induction. Adapun model yang terbentuk adalah seperti dibawah ini:



Gambar 2. Model T-Test pada naïve bayes, decision tree, k-NN, Deep Learning dan Rule Induction

Evaluasi

Perbandingan performance masing-masing algoritma diunjukkan dalam tabel 1 di bawah ini:

Tabel 1. Komparasi Accuracy dan AUC

	DT	NB	k-NN	RI	DL
Accura cy	96,50 %	73,82 %	96,50 %	89,87 %	99,00 %
AUC	0,866	0,786	0,814	0,776	0,997

Berdasarkan tabel 4 dapat diketahui bahwa algoritma Deep Learning memiliki nilai accuracy tertinggi yaitu 99,00, k-NN 96,50%, Decision Tree 96,50%, Rule Induction 89,87 dan naïve bayes 93,47%.

Sedangkan pada uji ROC curve menunjukkan bahwa Deep Learning mencapai nilai AUC yang terbaik yaitu 0,991, kemudian Decision Tree 0,866, k-NN 0,814, Naïve Bayes 0,786 dan Rule Induction 0,776.

Dalam pengujian ini, akan dibandingkan dua algoritma secara bergantian sehingga akan didapatkan hasil perbandingan seperti tabel 2:

Tabel 2. Uji statistik T-Test

A	B	C	D	E	F
	0.995 +/- 0.090	0.965 +/- 0.074	0.738 +/- 0.133	0.889 +/- 0.124	0.990 +/- 0.020
0.995 +/- 0.090		1.000	0.000	0.188	0.400
0.965 +/- 0.074			0.000	0.154	0.318
0.738 +/- 0.133				0.012	0.000
0.889 +/- 0.124					0.034
0.990 +/- 0.020					

Berdasarkan uji statistik T-Test pada tabel 2 di atas dapat dianalisis bahwa di atas dapat dianalisis bahwa perbedaan hasil t-test menunjukkan tidak terlalu signifikan perbedaan antar algoritmanya.

SIMPULAN DAN SARAN

Berdasarkan komparasi algoritma klasifikasi data mining yaitu Decision Tree, k-NN, Naïve Bayes, Deep Learning dan Rule Induction untuk prediksi status kontrak karyawan di sebuah perusahaan, maka dapat dianalisa hasil bahwa algoritma Deep Learning merupakan algoritma yang paling akurat. Sedangkan dalam pengujian terhadap nilai AUC Deep Learning masuk dalam kategori paling baik, Decision Tree dan k-NN masuk dalam kategori Baik, sedangkan Naïve Bayes dan Rule Induction masuk kedalam kategori Adil atau Sama. Dalam pengujian t-Test perbedaan hasil t-test menunjukkan tidak terlalu signifikan perbedaan antar algoritmanya.

DAFTAR PUSTAKA

Wahono, RS., Herman, NS., and Ahmad, S., A Comparison Framework of Classification Models for Software Defect Prediction, *Advanced Science Letters* Vol. 20, 1945-1950, 2014.

Yu, Lan., Chen, G., Koronios, A., Zhu, S., Guo, X., Application and Comparison of Classification Techniques in Controlling Credit Risk.

Quinlan, J. R. (1987). "Generating production rules from decision trees". In McDermott, John. *Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI-87)*. Milan, Italy. pp. 304-307.