

Prediksi Kinerja sebagai Rekomendasi Kenaikan Golongan dengan *Decision Tree* dan Regresi Logistik

<http://dx.doi.org/10.28932/jutisi.v8i1.4479>

Riwayat Artikel

Received: 04 Februari 2022 | Final Revision: 26 Februari 2022 | Accepted: 26 Februari 2022

Erik Dwi Anggara^{#1}, Andreas Widjaja^{✉#2}, Bernard Renaldy Suteja^{#3}

[#] Magister Ilmu Komputer, Universitas Kristen Maranatha
Jl. Prof. drg. Surya Sumantri No. 65, Bandung, Indonesia

¹2079004@maranatha.ac.id

²andreas.widjaja@it.maranatha.edu

³bernard.rs@it.maranatha.edu

Abstract — Employee performance is one element that greatly determines the quality of an organization, both government and private. Employee performance appraisal has become a routine for most companies. Performance appraisal is required for the process of salary increases, promotions, and demotions. Until this research was carried out, the processing of employee performance appraisal and evaluation at Prasama Bhakti Foundation was still done manually, so that sometimes employee promotions were carried out late or even on an inconsistent basis for each employee. Therefore, it is necessary to group data with the help of machine learning that can help predict the eligibility of an employee to get a promotion based on his performance. Classification is one method for classifying or classifying data that are arranged systematically. Decision tree and logistic regression methods are classification or grouping methods that have been widely used for solving classification problems. In this study, it will be explained how the process of processing employee performance appraisal data starts from data preparation to determine the accuracy of the decision tree model and logistic regression that is formed. The two classification models are used to predict employee performance as a recommendation for employee promotion at the Prasama Bhakti Foundation.

Keywords— decision; employee; performance; prediction; regression.

I. PENDAHULUAN

Kualitas suatu organisasi ditentukan oleh banyak sekali unsur. Salah satu unsur yang sangat menentukan adalah kinerja pegawai. Pegawai menjadi ujung tombak dari perusahaan, baik itu perusahaan jasa maupun produksi [1]. Pegawai pada lembaga pendidikan seperti sekolah, yaitu guru dan tenaga kependidikan merupakan pelayan konsumen sekolah yaitu peserta didik beserta orang tuanya

Pengukuran kinerja pegawai telah menjadi suatu rutinitas bagi sebagian besar perusahaan. Produktivitas sebuah perusahaan dapat ditingkatkan selain dari kinerja sistem keuangan, melainkan juga melalui pengukuran kinerja manusia (*human performance*) [2]. Pengukuran kinerja yang dilakukan oleh perusahaan dapat merangsang inisiatif dari pegawai untuk meningkatkan operasional kinerja, terutama ketika pegawai berpartisipasi secara mandiri dalam pengembangan ukuran kinerja departemen mereka sendiri [3]. Untuk mencapai kinerja yang unggul perlu dilakukan pengukuran kinerja masing-masing pegawai secara rinci. Kepercayaan dan tanggung jawab pegawai menjadi dasar mekanisme untuk mencapai keunggulan operasional [4]. Khusus bagi perusahaan pengelola produk jasa evaluasi kinerja dapat diterapkan berdasarkan perspektif pelanggan, selain yang dilakukan oleh pihak manajemen [5]. Berdasarkan pernyataan-pernyataan tersebut dapat disimpulkan bahwa pengukuran kinerja merupakan hal yang sangat penting, maka metode optimalisasi pengukuran kinerja terus dikembangkan, contohnya seperti konsep analisis maksimal dan tipikal kinerja yang dilakukan dari waktu ke waktu [6].

Penilaian kinerja pegawai diperlukan untuk proses kenaikan gaji, promosi jabatan, kenaikan dan penurunan pangkat dari pegawai yang bersangkutan. Sampai saat penelitian ini dilaksanakan, pengolahan penilaian dan evaluasi kinerja pegawai di yayasan Prasama Bhakti masih dilakukan dengan manual, sehingga terkadang promosi pegawai dilakukan terlambat atau

bahkan dengan dasar penilaian yang tidak konsisten untuk setiap pegawai. Oleh karena itu perlu dilakukan pengelompokan data dengan bantuan *machine learning* yang dapat membantu memprediksi kelayakan seorang pegawai mendapatkan promosi atau kenaikan pangkat berdasarkan kinerjanya.

Pengklasifikasian adalah salah satu cara atau metode yang digunakan untuk mengklasifikasikan atau mengelompokkan suatu data yang telah disusun secara sistematis. Salah satu metode pengelompokan atau pengklasifikasian yang banyak diterapkan dalam penelitian adalah metode *decision tree* atau pohon keputusan. Klasifikasi *decision tree* menyediakan metode untuk mengelompokkan *dataset* dengan cepat dan efektif [7]. Metode *decision tree* juga telah mengalami pengembangan untuk menghasilkan pendekatan yang dapat mengklasifikasi data-data yang bersifat sensitif [8]. Hasil klasifikasi dari pohon keputusan tersebut inilah yang kemudian digunakan untuk memprediksi kinerja dari seorang pegawai.

Selain *decision tree*, metode pengelompokan atau klasifikasi lain yang sering diterapkan adalah metode regresi logistik. Regresi logistik adalah suatu teknik statistika dalam menganalisis data yang bertujuan untuk mengetahui hubungan antara beberapa variabel dimana variabel respon yang bersifat kategorik, baik nominal maupun ordinal dengan variabel penjelas yang bersifat kategorik atau kontinu [9]. Saat variabel respon yang digunakan merupakan variabel dengan 2 macam kategori (dikotomis), maka regresi yang dilakukan menggunakan regresi logistik biner. Sedangkan pada saat variabel respon yang digunakan adalah variabel kategorik dengan lebih dari 2 macam kategori maka digunakan regresi logistik multinomial.

Beberapa penelitian sebelumnya juga telah dilakukan dengan tujuan yang sama yaitu menggunakan metode klasifikasi, baik dengan *decision tree* maupun dengan regresi logistik, untuk memprediksi suatu tujuan yang diinginkan. Penelitian pertama dilakukan oleh Isna Choina, Rachmat Aulia dan Ahmad Zakir [10]. Penelitian tersebut menggunakan pemrograman *Visual Basic 2010* dan menggunakan algoritma ID3 dalam menghasilkan *decision tree* dengan tujuan untuk dapat membantu perusahaan dalam menentukan keputusan dan memilih dengan tepat karyawan yang berprestasi.

Sejalan dengan penelitian sebelumnya, Chairani dan Syahputri R. mempublikasikan sebuah penelitian yang membahas tentang penerapan algoritma *Iterative Dichotomizer 3* (ID3) untuk menetapkan kelayakan perubahan status kerja karyawan (karyawan kontrak atau karyawan tetap) pada PT. Hanjung Indonesia. Karyawan kontrak PT. Hanjung Indonesia dikontrak selama satu tahun dan bagian kepegawaian akan menilai kinerja karyawan kontrak. Banyaknya jumlah karyawan kontrak, memungkinkan adanya ketidakakuratan dalam penyeleksian karyawan kontrak yang akan ditetapkan menjadi karyawan tetap. Selain itu penetapan perubahan status kerja dari karyawan kontrak menjadi karyawan tetap terkadang tidak dapat didefinisikan aturannya secara benar dan lengkap, karena data yang didapatkan tidak lengkap. Berdasarkan alasan tersebut diperlukan sebuah sistem cerdas berbasis komputer yang dapat memprediksi penetapan perubahan status kerja karyawan dari karyawan kontrak menjadi karyawan tetap. Kesimpulan yang dapat diambil dari penelitian tersebut adalah bahwa hasil *Learning* diperoleh nilai akurasi sebesar 91,30%. Pada tahap *post learning*, dari 41 data yang dilakukan proses prediksi diperoleh hasil sebanyak 23 karyawan akan naik status menjadi karyawan tetap, dan 18 karyawan akan tetap menjadi karyawan kontrak.

Penelitian ketiga dilakukan oleh Ninik Kristiyani, Andeka Rocky Tanaamah dan Charitas Fibriani [11]. Penelitian tersebut membuat sebuah sistem pendukung untuk mengambil keputusan yang dapat membantu perusahaan dalam menyeleksi calon karyawan. Di akhir penelitian dapat diperoleh kesimpulan bahwa pembuatan sistem pendukung keputusan dengan menerapkan algoritma ID3 merupakan hal yang dapat dilakukan. Sehingga sistem pendukung keputusan yang dihasilkan dapat dipakai oleh pihak HRD perusahaan dalam menentukan calon karyawan yang akan diterima dan untuk melaksanakan penilaian kinerja bagi karyawan.

Penelitian keempat menggunakan regresi logistik untuk melakukan prediksi. Penelitian tersebut dilakukan oleh R. Gunawan Santosa dan Antonius Rachmat Chrismanto [12]. Delapan model regresi logistik dihasilkan dalam penelitian ini untuk memprediksi prestasi mahasiswa. Berdasarkan akurasi model yang dihasilkan, model tersebut lebih baik digunakan untuk memprediksi jalur prestasi daripada jalur non prestasi.

Studi kasus lain berkenaan dengan klasifikasi dilakukan oleh Nursya'bani Hendro Prabowo dan Irhamah di PT. Gunung Meranti, Kalimantan Selatan [13]. Penelitian ini bertujuan membantu perusahaan menganalisa kinerja dari karyawan dalam menentukan target dan melakukan penilaian terhadap kinerja masing-masing karyawan dalam hal gaji pokok maupun bonus. Langkah yang dilakukan adalah dengan mengklasifikasikan kinerja karyawan berdasarkan penjualan produk perusahaan. Klasifikasi yang digunakan adalah metode Regresi Logistik dan *Naïve Bayes*. Hasil perbandingan ketepatan klasifikasi dan kebaikan model menunjukkan bahwa model yang dihasilkan dari Regresi Logistik maupun *Naïve Bayes* dapat dikategorikan baik.

Berdasarkan latar belakang tersebut dan didukung oleh sejumlah penelitian yang telah dilakukan sebelumnya, maka penelitian ini akan menggunakan metode klasifikasi *decision tree* dan regresi logistik dalam memprediksi kinerja pegawai sebagai rekomendasi kenaikan golongan ataupun promosi pegawai di Yayasan Prasama Bhakti. Dalam penelitian ini akan dijelaskan bagaimana proses pengolahan data penilaian kinerja pegawai mulai dari persiapan data hingga mengetahui keakuratan model *decision tree* dan regresi logistik yang terbentuk. Data yang digunakan dalam penelitian ini adalah data yang berasal dari Daftar Penilaian Pelaksanaan Pekerjaan (DP3) pegawai yang bekerja di Yayasan Prasama Bhakti. Sebagai

data tambahan, digunakan juga data *history* promosi pegawai yang telah dilakukan beserta data kepegawaian dan keterlambatan dari setiap pegawai.

II. METODE PENELITIAN

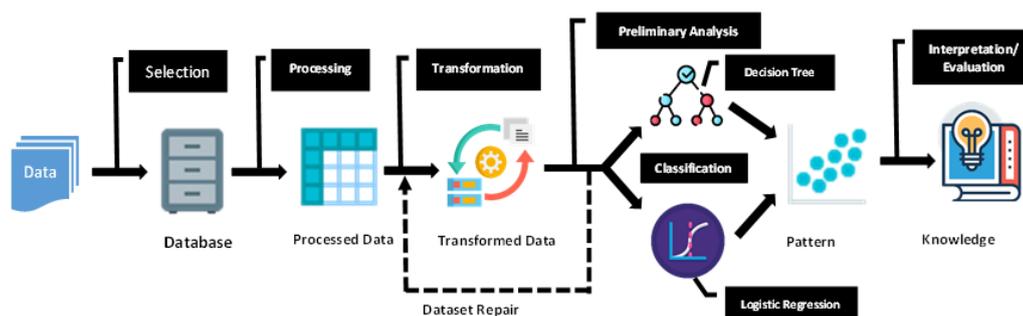
A. Data Mining

Penelitian kali ini akan mengikuti alur analisis *data mining*. *Data mining* atau penambangan data adalah proses pengumpulan informasi atau data penting yang diharapkan oleh suatu pihak melalui sebuah data yang lebih besar. Metode matematika, statistika, hingga teknologi *artificial intelligence* (AI) sering dimanfaatkan dalam proses pengumpulan informasi tersebut. *Data mining* bisa diartikan proses pencarian pola data yang tidak diketahui atau tidak diduga sebelumnya [14]. Ide dasarnya adalah menggali suatu sumber yang berharga dari tempat yang sama sekali tidak diperkirakan sebelumnya seperti perangkat lunak. Pola yang sebelumnya tidak tampak atau tidak begitu jelas diekstraksi menggunakan *data mining*.

Data mining memiliki banyak sekali fungsi. Namun fungsi yang paling utama adalah sebagai fungsi *descriptive* dan fungsi *predictive* [15]. Fungsi deskripsi dalam *data mining* adalah sebuah fungsi untuk memahami lebih jauh tentang data yang sedang diamati. Sedangkan fungsi prediksi adalah fungsi bagaimana sebuah proses nantinya akan menemukan pola tertentu dari suatu data. Pola-pola tersebut dapat diketahui dari berbagai variabel yang ada pada data. Pola yang sudah ditemukan tersebut dapat dipakai untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya. Oleh karena alasan tersebut, fungsi ini memudahkan dan menguntungkan bagi siapapun yang memerlukan prediksi yang akurat.

Data mining juga dikenal dengan istilah *knowledge discovery in databases* atau KDD. Banyak sekali teknik dan konsep yang dapat diaplikasikan pada aktivitas *data mining*. Agar dapat memperoleh data sesuai dengan yang diharapkan, proses tersebut memerlukan sejumlah langkah.

Proses *knowledge discovery* terdiri dari pembersihan data (*data cleaning*), integrasi data (*data integration*), pemilihan data (*data selection*), transformasi data (*data transformation*), *data mining*, evaluasi pola (*pattern evaluation*) dan penyajian pengetahuan (*knowledge presentation*) [16]. Tahapan proses KDD dalam *Data Mining* dapat dilihat pada gambar 1.



Gambar 1. Tahapan Proses KDD dalam *Data Mining*

1) Data Selection:

Proses untuk menyeleksi atau memilih data yang relevan atau berhubungan dengan analisis dan diterima dari kumpulan data yang sudah dikumpulkan sebelumnya. Sering kali tidak semua data yang ada pada *database* dapat langsung digunakan, hanya data yang sesuai untuk keperluan analisis yang akan diambil dari *database*. Data yang digunakan dalam penelitian ini adalah data yang berasal dari Daftar Penilaian Pelaksanaan Pekerjaan (DP3) pegawai yang bekerja di Yayasan Prasama Bhakti. DP3 sendiri adalah penilaian yang dilakukan oleh atasan kepada seorang pegawai bertujuan untuk memperoleh bahan-bahan pertimbangan yang obyektif dalam pendampingan pegawai binaan, dan dilaksanakan dalam kurun waktu sekali setahun oleh pejabat yang ditugaskan untuk menilai di yayasan ataupun unit sekolah yang berada di bawah yayasan tersebut. Dari DP3 tersebut diperoleh 864 row data yang berasal dari 216 pegawai dengan penilaian selama 4 tahun, mulai tahun 2017 hingga tahun 2020. Data yang diambil dari DP3 adalah kepribadian (K1), sosial (K2), pedagogi (K3), profesional (K4), dan tugas tambahan (K5). Hal tersebut tampak pada gambar 2.

3) Transformation Data:

Tahap *Transformation* dilakukan setelah tahap *Pre-Processing* dan *Cleaning Data*, yaitu tahap untuk menghasilkan data yang siap ditambah dengan mengubah data dari bentuk asalnya sesuai dengan kebutuhannya. Hal tersebut dilakukan agar dapat memudahkan proses penggalian data dalam menemukan suatu pengetahuan baru.

| No. Pegawai | Nama | Status | Lama Kerja (tahun) | Lama Kerja (hari) | K1 | K2 | K3 | K4 | K5 | Terlambat | Goal |
|-------------|------|--------|--------------------|-------------------|----------|----------|----------|----------|--------|-----------|------|
| | | 0 | 39 | 14366 | 92.65 | 85.69333 | 70 | 84.5 | 84.03 | 100 | 0 |
| | | 1 | 6 | 2435 | 88.20857 | 85.48667 | 82.66667 | 81.75 | 86.25 | 90.666667 | 1 |
| | | 1 | 7 | 2800 | 89.37857 | 78.75 | 83.72 | 80.08333 | 83.75 | 90.666667 | 0 |
| | | 1 | 2 | 1064 | 89.83857 | 84.44 | 86.88667 | 85 | 85.62 | 100 | 1 |
| | | 1 | 2 | 1064 | 87.11429 | 79.72 | 84.27667 | 82.16667 | 81.565 | 77.333333 | 1 |
| | | 0 | 2 | 883 | 84.40286 | 80.76 | 79.33333 | 81.58333 | 78.12 | 86.666667 | 0 |
| | | 1 | 2 | 883 | 87.46857 | 76.92667 | 70 | 78.04 | 70 | 100 | 1 |

Gambar 5. Dataset 1

Gambar 5 merupakan *dataset 1* yang siap untuk ditambah. Data dalam *dataset* tersebut berasal dari DP3, Daftar Hadir, Rekap Mutasi Kepangkatan, dan Daftar Urutan Kepangkatan. Setelah dilengkapi dengan data kunci berupa nomor pegawai, beberapa kolom data diambil dan dijadikan satu *dataset* berisikan data yang berkemungkinan besar menjadi atribut saat *data mining* dilakukan. *Dataset 1* berisikan kolom No. Pegawai, Nama, Status, Lama Kerja (tahun), Lama Kerja (hari), K1, K2, K3, K4, K5, Terlambat, dan Goal.

Kolom No. Pegawai berisikan data nomor pegawai dari pegawai yang bersangkutan. Nomor pegawai bersifat unik, berbeda untuk setiap pegawai. Kolom Nama berisikan nama lengkap dari pegawai. Kolom Status menyatakan status kepegawaian dari pegawai yang bersangkutan, yaitu pegawai tetap yayasan atau pegawai tidak tetap (honorar). Data tersebut kemudian diubah dalam bentuk *integer*, 0 untuk pegawai tidak tetap dan 1 untuk pegawai tetap yayasan. Lama kerja pegawai disajikan dalam dua bentuk di dua kolom berbeda, dalam tahun dan hari. Lama kerja dihitung mulai dari tanggal awal kerja pegawai hingga tanggal penilaian pegawai dibuat setiap tahunnya. K1 berisikan angka yang merupakan nilai akhir dari kriteria satu yaitu kepribadian. K2 berisikan angka yang merupakan nilai akhir dari kriteria dua yaitu sosial. K3 berisikan angka yang merupakan nilai akhir dari kriteria tiga yaitu tugas utama. K4 berisikan angka yang merupakan nilai akhir dari kriteria empat yaitu profesional. K5 berisikan angka yang merupakan nilai akhir dari kriteria lima yaitu tugas tambahan. Kolom terlambat berisikan jumlah keterlambatan pegawai yang bersangkutan selama satu tahun. Kolom terakhir Goal berisikan keputusan akhir dari penilaian kinerja pegawai tersebut. "Yes" berarti dipromosikan dan "No" berarti tidak dipromosikan. Data tersebut kemudian diubah menjadi data *integer*, 0 untuk "No" dan 1 untuk "Yes".

4) Classification

Proses menentukan tipe *data mining* yang akan digunakan dan memilih tugas *data mining* yang cocok. Pada tahap ini juga dilakukan pemilihan algoritma dan implementasi dari algoritma yang dipilih. Dua metode klasifikasi yang diterapkan dalam penelitian ini adalah metode klasifikasi menggunakan *Decision Tree* dan Regresi Logistik Biner. Kedua metode klasifikasi tersebut akan digunakan untuk memprediksi kinerja pegawai di Yayasan Prasama Bhakti. Alasan pemilihan metode *decision tree* dalam penelitian ini adalah konsep yang digunakan dalam metode *decision tree*, yaitu mengubah data yang dimiliki dalam tabel keputusan menjadi sebuah pohon keputusan dan aturan-aturan keputusan. Data tersebut berasal dari kriteria penilaian pegawai yang kemudian menjadi syarat dari layak tidaknya seorang pegawai mendapatkan promosi. Sedangkan regresi logistik biner dipilih karena variabel respon merupakan status kelayakan pegawai mendapatkan promosi, dimana variabel bernilai satu yang berarti pegawai tersebut layak mendapatkan promosi dan variabel bernilai nol yang berarti pegawai tersebut tidak layak mendapatkan promosi. Hasil klasifikasi kedua metode tersebut bersifat komplementer atau saling melengkapi, di mana kedua model tersebut sama-sama akan mengklasifikasikan atau menggolongkan hasil penilaian kinerja yang diperoleh oleh setiap pegawai layak digolongkan untuk mendapatkan promosi atau tidak. Sehingga pihak yayasan sendiri akan mendapatkan dua hasil prediksi yang dapat digunakan sebagai bahan rekomendasi dari keputusan yang akan diambil berkenaan dengan promosi setiap pegawai. Dalam pengerjaannya, proses *data mining* menggunakan Google Colaboratory sebagai *coding environment* dengan format *notebook* serupa dengan Jupyter Notebook gratis berbentuk *cloud* yang dijalankan menggunakan *browser*, seperti Mozilla Firefox dan Google Chrome.

5) Interpretation

Pada tahap ini pola hasil klasifikasi data yang diperoleh, dievaluasi atau diinterpretasikan dengan menggunakan *decision tree* dan regresi logistik. Pada tahap ini juga *data mining* memperoleh penggunaan dan umpan balik secara keseluruhan pada pola dan hasil penemuan.

6) Knowledge Presentation

Proses terakhir yang dimaksudkan untuk membantu pengguna dalam memahami serta menginterpretasikan hasil penambangan data menggunakan teknik visualisasi.

Tahap terakhir yang diperlukan dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Orang-orang yang tidak memahami *data mining* terkadang harus dilibatkan sehingga presentasi hasil *data mining* dalam berbentuk pengetahuan yang bisa dipahami semua orang. Visualisasi hasil juga dapat membantu mengkomunikasikan hasil data mini dalam presentasi ini.

B. Decision Tree

Decision tree adalah suatu metode klasifikasi menggunakan struktur pohon, dimana setiap *node* menerangkan atribut dan cabangnya menerangkan nilai dari atribut, sedangkan daunnya digunakan untuk menerangkan kelas. *Node* paling atas dari *decision tree* ini dinamakan *root*. *Decision tree* adalah salah satu metode yang dapat diterapkan untuk mengklasifikasikan tim objek atau data untuk menghasilkan sebuah keputusan [17].

Pendekatan ini terbagi dari serangkaian *node* pilihan, dihubungkan melalui cabang, bergerak menurun ke bawah dari simpul akar sampai berakhir di simpul daun. Pengembangan *decision tree* dimulai dari simpul akar, terutama berdasarkan pada konvensi yang diposisikan di bagian atas diagram pohon keputusan, semua atribut dievaluasi pada simpul seleksi, dengan setiap *outcome* yang mungkin menghasilkan cabang. Tiap cabang dapat masuk baik ke *decision node* yang lain ataupun ke *leaf node*.

Dalam pohon keputusan dikenal tiga jenis *node*, yaitu:

- 1) Akar sebagai *node* paling atas, *node* ini tidak memiliki input dan dapat tidak memiliki output atau dapat memiliki output lebih dari satu.
- 2) *Internal node* sebagai *node* percabangan, *node* ini hanya memiliki satu input dan output minimal dua.
- 3) Daun sebagai *node* akhir atau terminal *node*, *node* ini hanya memiliki satu input dan tidak memiliki output (simpul terminal).

Konsep dasar dari pohon keputusan adalah mengubah data menjadi model pohon keputusan, kemudian mengubah model pohon menjadi aturan (*rule*) dan menyederhanakannya. Manfaat utama dari menggunakan pohon keputusan adalah kemampuan untuk menyederhanakan proses pengambilan keputusan yang awalnya kompleks sehingga pembuat keputusan dapat menafsirkan solusi untuk masalah. Proses pembuatan *decision tree* menggunakan Google Colaboratory diawali dengan membuat perpustakaan yang diperlukan terlebih dahulu, kemudian dilanjutkan dengan membuat *Dataset 1* yang sudah siap untuk ditambang yang berasal dari file Excel.

Selanjutnya membagi kolom yang terdapat pada *Dataset 1* menjadi dua jenis variabel dependen (variabel target) dan variabel independen (variabel fitur). Variabel dependen atau target adalah data yang berasal dari kolom 'Goal'. Sedangkan variabel independen atau fitur yang digunakan berasal dari kolom "Lama Kerja", "K1", "K2", "K3", "K4", "K5" dan "Terlambat". Berikutnya *dataset* tersebut dibagi kembali menjadi *training set* dan *test set*. Besar *test set* yang digunakan adalah sebesar 10% dari data yang ada.

Langkah terakhir adalah membuat Model Pohon Keputusan menggunakan *Scikit-learn*. Pohon keputusan tersebut pun kemudian dievaluasi dengan mengetahui seberapa akurat pengklasifikasian dari model tersebut.

C. Algoritma Iterative Dichotomiser Three (ID3)

Algoritma *Iterative Dichotomiser Three* (ID3) merupakan algoritma yang paling dasar dari pembelajaran pohon keputusan yang melakukan pencarian secara menyeluruh pada semua kemungkinan *decision tree*.

Algoritma *Iterative Dichotomiser Three* (ID3) yang dikembangkan oleh J. Ross Quinlan merupakan salah satu algoritma induksi pohon keputusan. Algoritma ini dapat diterapkan menggunakan fungsi rekursif yaitu dengan memanggil dirinya sendiri. Algoritma ID3 berusaha membangun pohon keputusan dari atas ke bawah, yang mulai dengan pertanyaan : "atribut mana yang harus pertama kali di cek dan diletakan pada *root*". Pertanyaan tersebut dijawab dengan mengevaluasi semua atribut [18].

Langkah kerja metode ID3 adalah diawali dengan menentukan nilai bobot dari setiap atribut. Setelahnya dilanjutkan dengan proses pemilihan alternatif yang paling baik dari berbagai alternatif yang mungkin, dalam hal ini alternative yang dimaksud adalah dengan mengajukan alternatif yang berhak ditindaklanjuti berdasarkan kriteria-kriteria yang ditentukan. Proses tersebut akan terus berlanjut untuk proses yang sama (rekursif) dan nantinya akan terbentuk pohon keputusan. Apabila suatu atribut telah menjadi cabang (*node*) maka atribut tidak dimasukkan ke dalam perhitungan nilai *information gain*. Saat semua data dari anak cabang telah termasuk dalam kelas yang sama atau jika semua atribut telah digunakan tapi masih tersisa dalam kelas yang berbeda maka proses tersebut akan berhenti.

Rumus *Gain* dalam algoritma ID3 mengukur seberapa baik suatu atribut dalam memisahkan *training example* ke dalam kelas target. Atribut dengan informasi tertinggi akan dipilih. Pertama-tama digunakanlah ide dari teori informasi yang disebut *entropy* dengan tujuan untuk mendefinisikan *gain*. *Entropy* mengukur jumlah dari informasi yang ada pada atribut.

Entropy digunakan untuk menghitung nilai keberagaman dari suatu kumpulan sampel data. Semakin besar nilai *entropy*-nya berarti semakin heterogen kumpulan data tersebut. Begitu juga sebaliknya. *Entropy* merupakan jumlah bit yang dibutuhkan untuk menyatakan suatu kelas, baik positif (+) maupun negatif (-) dari ketidakpastian yang berdasar pada variabel acak pada jumlah data sampel. Panjang kode untuk menyatakan informasi secara optimal adalah $-\log_2 p$ bit untuk *messages* yang mempunyai probabilitas p [18].

Rumus untuk menghitung *entropy* sebagai berikut [19]:

$$Entropy(S) = -P(+)\log_2 P(+)-P(-)\log_2 P(-) \quad (1)$$

Dimana :

- S = banyaknya sampel data yang digunakan untuk *training*.
- $P(+)$ = jumlah probabilitas bersolusi positif (mendukung) pada data sampel untuk kriteria tertentu
- $P(-)$ = jumlah probabilitas bersolusi negatif (tidak mendukung) dari data sampel untuk kriteria tertentu.

Efektivitas suatu atribut dari suatu klasifikasi data dapat diukur setelah mendapatkan *Entropy* dari suatu kumpulan data sampel. Ukuran efektivitas suatu atribut inilah yang disebut dengan *information gain*. *Information gain* dari suatu atribut dapat dihitung dengan menggunakan persamaan berikut [19]:

$$Gain(S, A) = Entropy(S) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} entropy(S_v) \quad (2)$$

Dimana:

- A = Atribut
- v = Nilai yang mungkin untuk atribut A
- $values(A)$ = Himpunan yang mungkin untuk atribut A
- $|S_v|$ = Jumlah sampel dari v
- $|S|$ = Jumlah sampel seluruh sampel data
- $entropy(S_v)$ = Entropy untuk sampel-sampel v

D. Regresi Logistik

Regresi logistik dalam statistika merupakan suatu teknik analisis data yang bertujuan untuk mengetahui hubungan antara beberapa variabel dimana variabel responnya bersifat kategorik, baik nominal maupun ordinal dengan variabel penjelasnya bersifat kategorik atau kontinu [9].

Secara khusus regresi logistik yang digunakan pada penelitian ini adalah regresi logistik biner. Regresi logistik biner merupakan salah satu pendekatan model matematis yang digunakan untuk menganalisis hubungan beberapa faktor dengan sebuah variabel yang bersifat biner, yaitu variabel respon yang terdiri dari dua kategori misalnya $y = 1$ menyatakan hasil yang diperoleh "sukses" dan $y = 0$ menyatakan hasil yang diperoleh "gagal" [20].

Dalam kondisi demikian, variabel y mengikuti distribusi Bernoulli untuk setiap observasi tunggal. Fungsi Probabilitas untuk setiap observasi adalah diberikan sebagai berikut [21].

$$f(y) = \pi^y(1 - \pi)^{1-y}; \quad y = 0, 1 \quad (3)$$

Suatu fungsi $P(X|Y = 1)$ dicari dengan menggunakan transformasi logit yaitu $g(x)$ dinyatakan sebagai berikut [21]:

$$P(X|Y = 1) = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}} \quad (4)$$

Dimana:

- $P(X|Y = 1)$: probabilitas kejadian berhasil
- p : banyaknya variabel prediktor
- $X_1 \dots$: Variabel independen kuantitatif atau kualitatif
- β_0 : Konstanta
- $\beta_1 \dots \beta_p$: Koefisien variabel regresi

Berdasarkan persamaan (4), p adalah banyaknya variabel prediktor. Untuk mempermudah pendugaan parameter regresi, $P(X|Y = 1)$ merupakan peluang kejadian sukses dengan nilai probabilitas $0 \leq P(X|Y = 1) \leq 1$ dan merupakan nilai parameter untuk $j = 1, 2, \dots, p$. $P(X|Y = 1)$ merupakan fungsi yang non linear, sehingga perlu ditransformasi ke dalam bentuk logit untuk memperoleh fungsi yang linear agar hubungan antara variabel independen dan variabel dependen dapat terlihat [22]. Model transformasi logit dari $\pi(x)$ dari persamaan (4) dapat dituliskan sebagai berikut [21]:

$$g(x) = \ln \left(\frac{P(X|Y = 1)}{1 - P(X|Y = 1)} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (5)$$

$g(x)$: transformasi logit dari $P(X|Y = 1)$

Model tersebut merupakan fungsi linier dari parameter-parameternya. Dalam model regresi linier, diasumsikan bahwa amatan dari variabel respon diekspresikan sebagai $y = E(Y|x) + \varepsilon$ terdapat pada persamaan (6)

$$E(Y|x) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (6)$$

merupakan rata-rata dari populasi dan ε merupakan komponen acak yang menunjukkan penyimpangan amatan dari rata-ratanya dan ε diasumsikan mengikuti sebaran normal dengan rata-rata nol dan varians konstan.

Untuk mengembalikan nilai probabilitas yang dihasilkan kepada dua kategori variabel responnya, yaitu $y = 0$ atau $y = 1$ digunakan konversi sebagai berikut:

$$\hat{y}(x) = \begin{cases} 0, & P(X|Y = 1) < 0.5 \\ 1, & P(X|Y = 1) \geq 0.5 \end{cases}$$

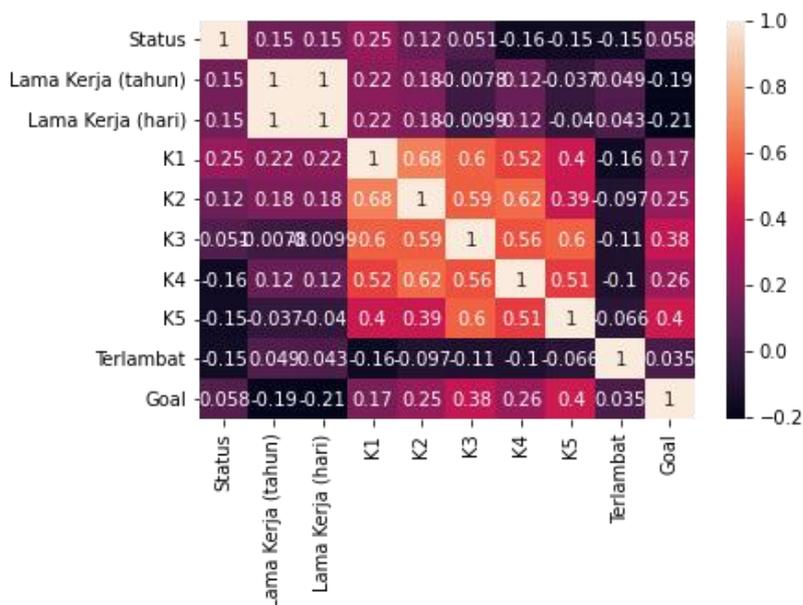
Proses pembuatan model regresi logistik juga menggunakan Google Colaboratory sebagai *coding environment* dengan format *notebook* serupa dengan Jupyter.

III. HASIL DAN PEMBAHASAN

A. Analisis Data Awal

Seperti yang telah diuraikan pada bagian sebelumnya, untuk teknik analisis data yang digunakan pada penelitian ini adalah klasifikasi data dengan *decision tree* dan regresi logistik. *Dataset 1* yang dimuat ke dalam Google Colaboratory berisikan kolom No. Pegawai, Nama, Status, Lama Kerja (tahun), Lama Kerja (hari), K1, K2, K3, K4, K5, Terlambat, dan Goal. Tidak semua kolom tersebut akan digunakan sebagai fitur dalam melakukan klasifikasi. Demikian halnya dengan kolom No. Pegawai dan Nama. Kedua kolom tersebut dibuang dari awal karena prediksi kinerja yang dilakukan tidak bersifat subjektif. Prediksi yang dilakukan hanya melihat dari hasil penilaian dari setiap pegawai.

Sisa kolom yang ada kemudian diinvestigasi seluruh hubungan antar variabel numerik dalam *dataset* dengan mencari nilai korelasinya (r). Matriks Korelasi dapat dibuat dengan menggunakan fungsi *corr()*. Fungsi *corr()* yang digunakan dalam Python secara *default* memiliki parameter *method* bernilai Pearson, sehingga korelasi yang digunakan adalah korelasi Pearson. Korelasi Pearson merupakan salah satu teknik analisis yang berguna untuk mengukur kuat lemahnya hubungan antara dua variabel. Nilai korelasi (r) yang dihasilkan dapat bernilai positif atau negatif dan berada pada rentang -1 dan 1 . Apabila nilai r mendekati -1 atau 1 maka hubungan kedua variabel semakin kuat, sebaliknya jika nilai r mendekati 0 maka hubungan kedua variabel semakin lemah. Matriks Korelasi tersebut kemudian divisualisasikan kemudian dievaluasi untuk melihat bagaimana hubungan setiap variabel yang ada terutama dengan variabel target "Goal".



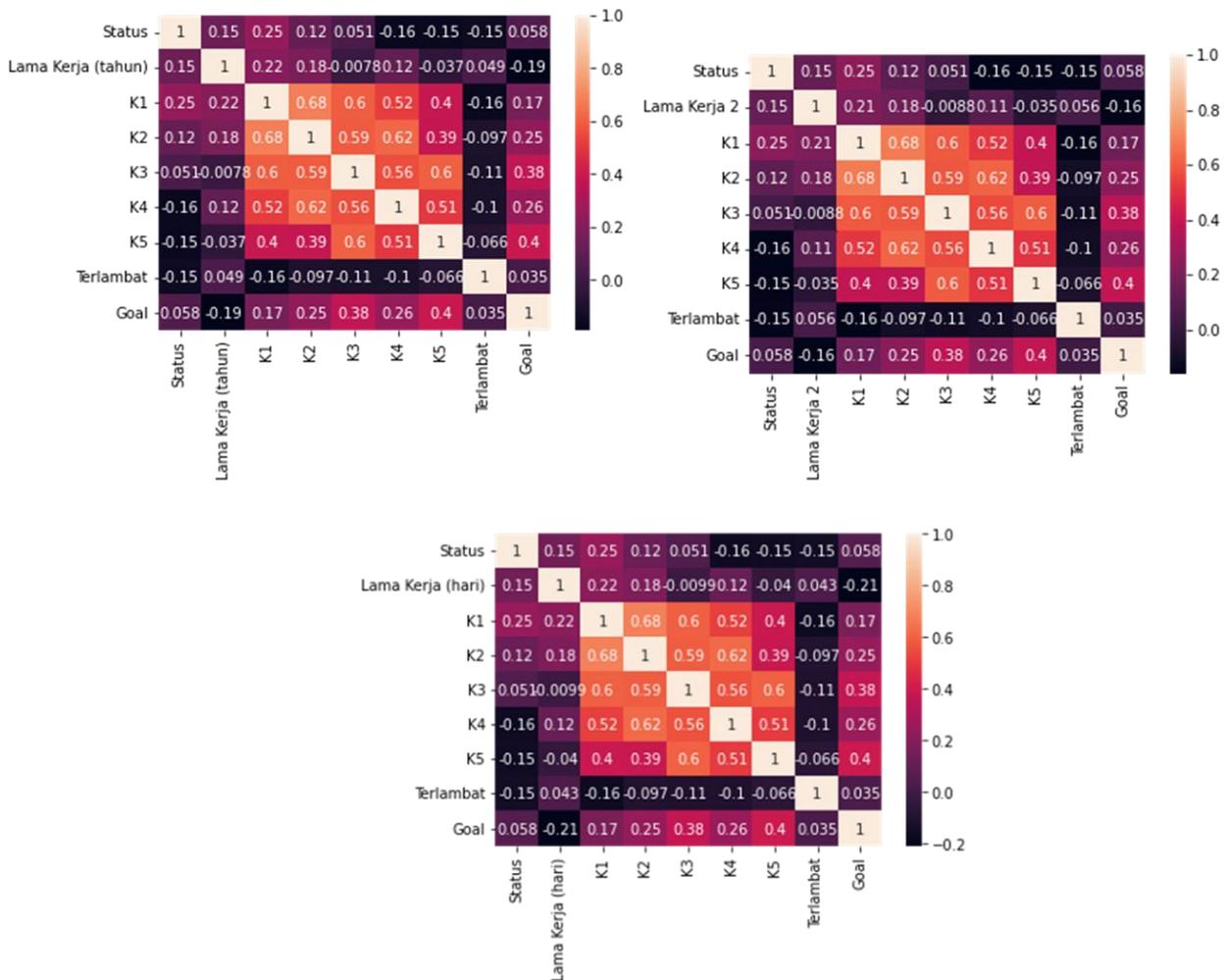
Gambar 6. Matriks Korelasi Dataset 1

Berdasarkan gambar 6, terlihat bahwa setiap variabel pada Dataset 1 menunjukkan adanya korelasi dengan variabel target “Goal”. Walaupun memang korelasi yang ditunjukkan bukan korelasi yang kuat. Terdapat satu variabel yang menunjukkan korelasi lemah dengan variabel target, yaitu variabel “Lama Kerja”. Hal ini terlihat dengan perolehan nilai korelasi yang mendekati nilai nol.

Menurut ketentuan yang ada di Yayasan Prasama Bhakti, lama kerja pegawai di yayasan seharusnya memiliki korelasi dengan keputusan akhir, apakah pegawai tersebut akan dipromosikan atau tidak. Kenaikan pangkat atau promosi merupakan kesempatan yang diberikan kepada pegawai setiap dua tahun sekali setelah memenuhi semua kondisi yang disyaratkan. Hanya saja seperti yang dijelaskan sebelumnya, penilaian yang masih dilakukan manual membuat keputusan yang diambil kurang tepat atau tidak sesuai. Hal ini menandakan bahwa lama kerja pegawai menjadi penentu awal kesempatan promosi diberikan. Oleh karena itu diputuskan untuk mencoba mentransformasi kembali data lama kerja tersebut ke dalam beberapa kemungkinan bentuk lain yang dapat meningkatkan korelasinya terhadap variabel target.

Kemungkinan pertama yang dilakukan adalah dengan memisahkan lama kerja dalam satuan tahun dengan lama kerja dalam satuan hari. Kedua variabel tersebut sebenarnya adalah data yang sama. Hal ini terlihat pada hasil korelasi antara kedua variabel tersebut memiliki korelasi yang pasti seperti yang divisualisasikan pada gambar 6.

Dataset 2 dibentuk dengan menyisakan lama kerja dalam tahun. Kolom tersebut berisikan data lama kerja pegawai yang telah dibulatkan ke bawah. Dataset 3 merupakan Dataset 2 yang mengalami perubahan pada kolom lama kerja. Data pada kolom lama kerja dalam dataset 3 merupakan lama kerja pegawai yang dinyatakan dalam kelipatan dua. Lama kerja dalam tahun dibulatkan ke bawah menuju angka kelipatan dua terdekat. Dataset 4 merupakan hasil pemisahan Dataset 1, dimana lama kerja yang terdapat pada dataset ini dalam satuan hari. Dataset 2, 3, dan 4 diujikan kembali melalui Matriks Korelasi sebagai berikut.



Gambar 7. Matriks Korelasi *Dataset 2, 3, dan 4.*

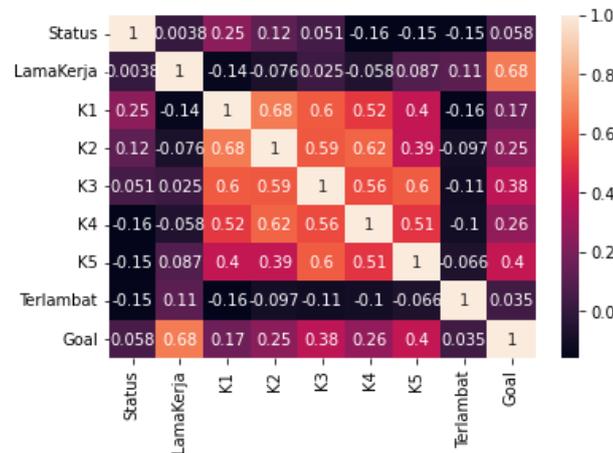
Gambar 7 adalah hasil visualisasi matriks korelasi untuk *dataset 2, 3 dan 4.* Terlihat pada Matriks Korelasi tidak ada perubahan korelasi dari variabel lama kerja terhadap variabel target “Goal”. Baik lama kerja dalam tahun asli, tahun kelipatan dua, maupun dalam hari belum menunjukkan adanya korelasi yang diharapkan. Oleh karena itu, perubahan terhadap data lama kerja pegawai masih perlu dilakukan agar dapat memperoleh korelasi antara lama kerja pegawai dengan variabel target.

| | Status | LamaKerja | K1 | K2 | K3 | K4 | K5 | Terlambat | Goal |
|---|--------|-----------|-----------|-----------|-----------|-----------|--------|------------|------|
| 0 | 0 | 0 | 92.650000 | 85.693333 | 70.000000 | 84.500000 | 84.030 | 100.000000 | 0 |
| 1 | 1 | 1 | 88.208571 | 85.486667 | 82.666667 | 81.750000 | 86.250 | 90.666667 | 1 |
| 2 | 1 | 0 | 89.378571 | 78.750000 | 83.720000 | 80.083333 | 83.750 | 90.666667 | 0 |
| 3 | 1 | 1 | 89.838571 | 84.440000 | 86.886667 | 85.000000 | 85.620 | 100.000000 | 1 |
| 4 | 1 | 1 | 87.114286 | 79.720000 | 84.276667 | 82.166667 | 81.565 | 77.333333 | 1 |

Gambar 8. *Dataset 5.*

Dataset 5 pada gambar 8 berasal dari *Dataset 2* dengan data lama kerja dikelompokkan menjadi ganjil dan genap. Langkah ini diambil karena aturan yayasan sendiri yang memberikan kesempatan kepada pegawai untuk mendapatkan kenaikan

golongan ataupun promosi lainnya setiap dua tahun sekali. Sehingga data lama kerja pegawai yang sebelumnya menggunakan satuan tahun diubah menjadi data integer dengan dua kategori saja, yaitu “0” untuk genap dan “1” untuk ganjil. Lama kerja genap berarti pegawai tersebut mempunyai lama kerja kelipatan dua dan mendapatkan kesempatan promosi, sedangkan lama kerja ganjil berarti pegawai tersebut lama kerja bukan kelipatan dua dan tidak mendapatkan kesempatan promosi. Sama seperti sebelumnya *Dataset 5* tersebut kemudian diujikan kembali melalui Matriks Korelasi sebagai berikut:



Gambar 9. Matriks Korelasi *Dataset 5*

Hasil visualisasi Matriks Korelasi pada gambar 9 dari *Dataset 5* menunjukkan adanya korelasi antara lama kerja pegawai yang telah digolongkan ke dalam angka ganjil genap dengan variabel target “Goal”. Terlihat perubahan yang cukup signifikan dengan warna yang kini mendekati nilai korelasi sangat tinggi.

Dari hasil tersebut dapat disimpulkan bahwa data lama kerja yang telah digolongkan ke dalam ganjil genap dapat dikatakan cukup banyak mempengaruhi layak tidaknya seorang pegawai mendapatkan kenaikan golongan atau promosi. Hal ini sesuai dengan aturan yang berlaku di Yayasan Prasama Bhakti. Dengan menggolongkan lama kerja ke dalam angka ganjil genap memudahkan *machine learning* untuk menangkap pola kriteria promosi berdasarkan lama kerja pegawai, *Dataset* sebelumnya dengan data lama kerja pegawai dalam tahun maupun hari sulit untuk direkam polanya, dikarenakan *range* lamanya pegawai di Yayasan Prasama Bhakti yang bervariasi, mulai dari pegawai yang baru masuk hingga pegawai senior yang mendekati masa pensiun. Ditambah lagi dengan jumlah data yang digunakan dalam *dataset* yang tidak banyak, semakin menyulitkan untuk mendapatkan kriteria berdasarkan lama kerja dengan angka yang sama.

Setelah melihat hasil dari lima kali uji coba perubahan data pada variabel lama kerja, maka diputuskan untuk menggunakan *Dataset 5* sebagai *dataset* utama yang akan digunakan dalam memprediksi kinerja pegawai dengan menggunakan *decision tree* dan regresi logistik.

B. Analisis Decision Tree

Berdasarkan hasil uji coba dari beberapa bentuk *dataset* di atas, diambil keputusan untuk menggunakan *Dataset 5* dalam membentuk *Decision Tree*. Setelah memuat seluruh *libraries* yang dibutuhkan dan memuat *Dataset 5* langkah berikut yang dilakukan adalah menyeleksi fitur dengan membagi kolom yang ada pada *dataset* menjadi dua jenis variabel, yaitu variabel dependen atau variabel target dan variabel independen atau variabel fitur. Variabel independen atau fitur yang digunakan berasal dari kolom “LamaKerja”, “K1”, “K2”, “K3”, “K4”, “K5” dan “Terlambat”, sedangkan variabel dependen atau target berasal dari kolom “Goal”.

Selanjutnya untuk memahami kinerja model, *dataset* tersebut dibagi menjadi *training set* dan *test set*. *Dataset* dipisahkan menggunakan fungsi *train_test_split()*. Besar *test set* yang digunakan adalah sebesar 10% dari data yang ada.

Langkah berikutnya adalah membuat Model Pohon Keputusan menggunakan *Scikit-learn* menggunakan fungsi *DecisionTreeClassifier()* yang diterapkan pada *training set*. Model tersebut juga divisualisasikan agar dapat terlihat bagaimana *decision tree* yang terbentuk. Hasil visualisasi *decision tree* dapat dilihat pada halaman lampiran. Dari hasil visualisasi tampak bahwa atribut atau variabel yang pertama digunakan sebagai penentu keputusan adalah lama kerja. Hal ini diketahui dari variabel lama kerja yang menjadi akar atau *node* paling atas dari *decision tree*. Nilai variabel lama kerja dilihat apakah nilainya kurang atau lebih dari 0,5 sesuai dengan penggolongan lama kerja genap (0) atau lama kerja ganjil (1). Atribut berikutnya seimbang antara K1, K2, K3, K4, dan K5 yang secara bergantian digunakan untuk mencari nilai kriteria

dari kelima variabel tersebut. Setelah semua data selesai dikelompokkan berdasarkan atribut-atribut tersebut, data yang tersisa di akhir barulah dipisahkan berdasarkan jumlah keterlambatan.

Hasil prediksi dari *decision tree* yang terbentuk juga dapat diujikan dengan menginput data riil secara berurutan dari lama kerja pegawai (ganjil/genap), nilai K1, K2, K3, K4, K5 dan jumlah keterlambatan pegawai yang bersangkutan, seperti pada gambar 10.

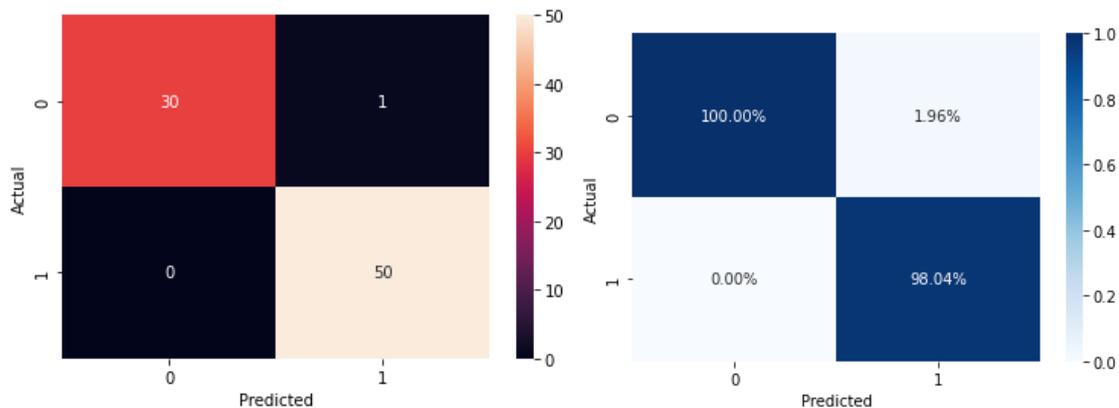
```
print(dtree.predict([[1, 60, 62, 64, 62, 65, 0]]))  
[0]
```

Gambar 10. Contoh Prediksi dengan Menggunakan Model *Decision Tree*.

Contoh prediksi pada gambar 10 menghasilkan prediksi bernilai 0, yang berarti pegawai dengan data yang diinput tersebut tidak direkomendasikan untuk mendapatkan promosi.

Setelah model *decision tree* terbentuk dilanjutkan dengan menguji model tersebut untuk memprediksi hasil dari data tes. Kumpulan hasil prediksi tersebut dinyatakan sebagai y_{pred} .

Tahap berikutnya adalah mengevaluasi model tersebut. Untuk mengetahui seberapa baik model yang terbentuk perlu dibuat *confusion matrix* yang membandingkan hasil prediksi yang dilakukan oleh model tersebut dengan hasil sebenarnya dari data tes.



Gambar 11. *Confusion Matrix* dari *Decision Tree* dalam Jumlah (kiri) dan Persentase (kanan).

Dari *confusion matrix* pada gambar 11 diketahui bahwa model *decision tree* berhasil memprediksi 100% data negatif atau sebanyak 30 data dan 98,04% data positif atau sebanyak 50 data. Selanjutnya untuk lebih memastikan tingkat akurasinya, evaluasi model juga dilakukan dengan mengetahui seberapa besar nilai akurasi model dengan membandingkan data pada *test set* dan data hasil prediksi. Dari hasil yang diperoleh pada gambar *confusion matrix* dapat dihitung nilai akurasi dari model *decision tree* yang terbentuk dengan membandingkan jumlah data berhasil diprediksi, baik data positif maupun negatif, yaitu sebanyak 80 data, dengan seluruh jumlah data yang ada pada *test set*, yaitu sebanyak 81 data. Hasil perhitungan tersebut sesuai dengan perhitungan Google Colaboratory menggunakan fungsi *accuracy_score()* dengan perolehan nilai akurasi sebesar $\frac{80}{81} \cong 0.9876$ untuk model *decision tree* yang terbentuk.

Sebagai pembandingan dilakukan juga proses sama dalam pembentukan model *decision tree* dengan memisahkan *test set* sebesar 20% dari *dataset* yang sama. Dari proses *test* data tersebut menghasilkan nilai akurasi dari model *decision tree* yang terbentuk sebesar 0.9383.

C. Analisis Regresi Logistik

Melanjutkan proses *data mining* sebelumnya dengan menggunakan *decision tree*, proses klasifikasi regresi logistik dilakukan pada *Dataset 5* yang telah diseleksi fitur dan telah dipisahkan antara *training set* dan *test set*. Setelah mengimpor modul *Logistic Regression* dibuatlah objek *classifier Logistic Regression* menggunakan fungsi *LogisticRegression()*. Model yang terbentuk kemudian digunakan untuk menghasilkan prediksi dari *test set*.

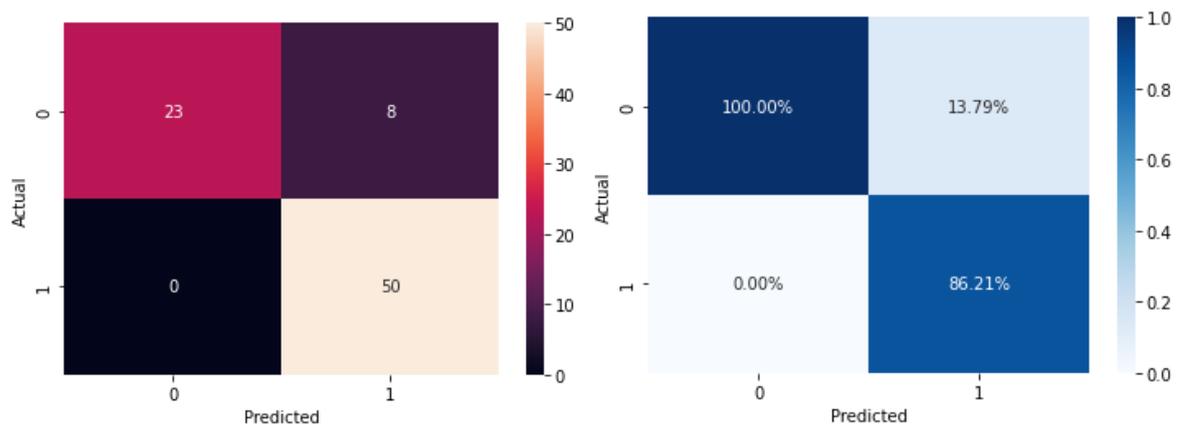
Hasil prediksi dari model regresi logistik yang terbentuk juga dapat diujikan dengan menginput data riil secara berurutan dari lama kerja pegawai (ganjil/genap), nilai K1, K2, K3, K4, K5 dan jumlah keterlambatan pegawai yang bersangkutan, seperti pada gambar 12.

```
print(model.predict([[1, 60, 62, 64, 62, 65, 0]]))  
[0]
```

Gambar 12. Contoh Prediksi dengan Menggunakan Model Regresi Logistik.

Contoh prediksi pada gambar 12 menghasilkan prediksi bernilai 0, yang berarti pegawai dengan data yang diinput tersebut tidak direkomendasikan untuk mendapatkan promosi.

Setelah model regresi logistik terbentuk dan prediksi dihasilkan, tahap berikutnya adalah mengevaluasi model tersebut. Sejalan dengan *decision tree*, *confusion matrix* dibuat untuk mengetahui seberapa baik model regresi logistik yang terbentuk.



Gambar 13. Confusion Matrix dari Regresi Logistik dalam Jumlah (kiri) dan Persentase (kanan).

Dari *confusion matrix* pada gambar 13 diketahui bahwa model regresi logistik berhasil memprediksi 100% data negatif atau sebanyak 23 data dan 86,21% data positif atau sebanyak 50 data.

Selanjutnya untuk lebih memastikan tingkat akurasi, evaluasi model juga dilakukan dengan mengetahui seberapa besar nilai akurasi model dengan membandingkan data pada *test set* dan data hasil prediksi. Dari hasil yang diperoleh pada gambar *confusion matrix* dapat dihitung nilai akurasi dari model *decision tree* yang terbentuk dengan membandingkan jumlah data berhasil diprediksi, baik data positif maupun negatif, yaitu sebanyak 73 data, dengan seluruh jumlah data yang ada pada *test set*, yaitu sebanyak 81 data. Hasil perhitungan tersebut sesuai dengan perhitungan Google Colaboratory menggunakan fungsi *accuracy_score()* dengan perolehan nilai akurasi sebesar $\frac{73}{81} \cong 0.9012$ untuk model regresi logistik yang terbentuk.

Sebagai pembandingan dilakukan juga proses sama dalam pembentukan model regresi logistik dengan memisahkan *test set* sebesar 20% dari *dataset* yang sama. Dari proses *test* data tersebut menghasilkan nilai akurasi dari model regresi logistik yang terbentuk sebesar 0.9074.

Tahap akhir dalam analisis regresi logistik adalah pembuatan persamaan dari regresi logistik yang nantinya dapat digunakan untuk memprediksi hasil kinerja pegawai, dengan menginput nilai atribut yang berasal dari penilaian pegawai ke dalam persamaannya. Untuk membentuk persamaan regresi logistik itu, perlu dicari nilai koefisien setiap variabel dan konstanta dari persamaan regresi logistik.

```
import statsmodels.formula.api as smf
logit_model=smf.logit("Goal ~ LamaKerja + K1 + K2 + K3 + K4 + K5 + Terlambat", data=df)
result=logit_model.fit()
print(result.summary())
```

Optimization terminated successfully.
Current function value: 0.278124
Iterations 7

Logit Regression Results

```
=====
Dep. Variable:          Goal    No. Observations:          810
Model:                 Logit    Df Residuals:              802
Method:                MLE     Df Model:                  7
Date:                  Thu, 27 Jan 2022    Pseudo R-squ.:            0.5670
Time:                  05:59:57    Log-Likelihood:           -225.28
converged:              True     LL-Null:                  -520.29
Covariance Type:      nonrobust    LLR p-value:              3.423e-123
=====
```

| | coef | std err | z | P> z | [0.025 | 0.975] |
|-----------|----------|---------|--------|-------|---------|---------|
| Intercept | -16.9410 | 1.990 | -8.513 | 0.000 | -20.841 | -13.041 |
| LamaKerja | 4.9105 | 0.342 | 14.358 | 0.000 | 4.240 | 5.581 |
| K1 | -0.0179 | 0.019 | -0.924 | 0.355 | -0.056 | 0.020 |
| K2 | 0.0290 | 0.022 | 1.349 | 0.177 | -0.013 | 0.071 |
| K3 | 0.0789 | 0.018 | 4.328 | 0.000 | 0.043 | 0.115 |
| K4 | 0.0085 | 0.019 | 0.450 | 0.653 | -0.028 | 0.045 |
| K5 | 0.0898 | 0.018 | 4.899 | 0.000 | 0.054 | 0.126 |
| Terlambat | -0.0032 | 0.012 | -0.262 | 0.793 | -0.027 | 0.021 |

Gambar 14. Tampilan Nilai Konstanta dan Koefisien Persamaan Regresi Logistik

Dari gambar 14 diketahui bahwa nilai konstanta (Intercept) adalah -16,9410, koefisien LamaKerja adalah 4,9105, koefisien K1 adalah - 0,0179, koefisien K2 adalah 0,0290, koefisien K3 adalah 0,0789, koefisien K4 adalah 0,0085, koefisien K5 adalah 0,0898, dan koefisien Terlambat adalah - 0. 0032.

Hasil yang sama juga ditunjukkan hasil perhitungan regresi logistik dengan menggunakan SPSS untuk data yang sama.

Variables in the Equation

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. for EXP(B) | |
|---------------------|-----------|---------|-------|---------|----|------|---------|---------------------|---------|
| | | | | | | | | Lower | Upper |
| Step 1 ^a | LamaKerja | 4.911 | .342 | 206.156 | 1 | .000 | 135.708 | 69.421 | 265.288 |
| | K1 | -.018 | .019 | .854 | 1 | .355 | .982 | .946 | 1.020 |
| | K2 | .029 | .022 | 1.821 | 1 | .177 | 1.029 | .987 | 1.074 |
| | K3 | .079 | .018 | 18.729 | 1 | .000 | 1.082 | 1.044 | 1.122 |
| | K4 | .008 | .019 | .203 | 1 | .653 | 1.009 | .972 | 1.046 |
| | K5 | .090 | .018 | 24.000 | 1 | .000 | 1.094 | 1.055 | 1.134 |
| | Terlambat | -.003 | .012 | .069 | 1 | .793 | .997 | .973 | 1.021 |
| | Constant | -16.941 | 1.990 | 72.468 | 1 | .000 | .000 | | |

a. Variable(s) entered on step 1: LamaKerja, K1, K2, K3, K4, K5, Terlambat.

Gambar 15. Tampilan Nilai Konstanta dan Koefisien Persamaan Regresi Logistik dengan SPSS

Terlihat pada gambar 15 nilai konstanta dan koefisien dari persamaan regresi logistik yang terbentuk menghasilkan nilai yang sama dengan menggunakan taraf signifikansi $\sigma = 5\%$. Walaupun ada sedikit perbedaan dikarenakan hasil dari SPSS dibulatkan ke dalam angka dengan tiga digit di belakang tanda koma. Sehingga diperoleh persamaan logistik sebagai berikut:

$$P(X|Y = 1) = \frac{1}{1 + e^{-(-16,941+4,911X_1-0,018X_2+0,029X_3+0,079X_4+0,008X_5+0,090X_6-0,003X_7)}} \quad (6)$$

Dimana:

- $P(X|Y = 1)$: probabilitas promosi atau kenaikan golongan
 X_1 : nilai variabel lama kerja
 X_2 : nilai variabel K1
 X_3 : nilai variabel K2
 X_4 : nilai variabel K3
 X_5 : nilai variabel K4
 X_6 : nilai variabel K5
 X_7 : nilai variabel Terlambat

Dari hasil perhitungan SPSS gambar 15 kolom Sig menginformasikan signifikan pengaruh variabel independen terhadap variabel dependen. Terjadi pengaruh yang signifikan jika nilai Sig < 0,05. Tampak variabel yang berpengaruh terhadap peluang promosi atau kenaikan golongan pegawai adalah “LamaKerja”, “K3”, dan “K5” dengan nilai Sig yang sama. Sedangkan untuk variabel “K1”, “K2”, “K4” dan “Terlambat” tidak berpengaruh terhadap peluang seorang pegawai mendapatkan promosi atau kenaikan golongan. Nilai yang sama juga dapat dilihat pada gambar 14, tampilan dari Google Colaboratory, pada kolom “ $P > |z|$ ”.

Berdasarkan nilai Sig dari setiap variabel yang digunakan sebagai fitur dalam metode regresi logistik, dapat diambil kesimpulan bahwa kriteria-kriteria penilaian pegawai yang dilakukan di Yayasan Prasama Bhakti tidak semuanya berpengaruh terhadap keputusan seorang pegawai memperoleh kesempatan kenaikan golongan atau promosi. Kriteria dalam DP3 dalam hal ini pedagogi (K3) dan tugas tambahan (K5) menjadi kriteria yang paling berpengaruh dalam penilaian. Hal ini menandakan bahwa kemampuan seorang pegawai dalam merencanakan, melaksanakan, dan evaluasi kegiatan pembelajaran menjadi hal yang paling penting dalam menentukan pegawai tersebut mendapatkan hasil penilaian yang baik. Tentu saja ini merupakan hal yang baik, mengingat Yayasan Prasama Bhakti merupakan yayasan yang bergerak di bidang pendidikan yang sebagian besar pegawainya adalah guru. Kemampuan pegawai dalam menjalankan tugas tambahannya pun ternyata menjadi kriteria yang berpengaruh dalam penilaian. Hal ini menandakan pegawai yang mempunyai tugas tambahan diluar tugas utamanya dan mampu menjalankannya dengan baik mempunyai nilai yang lebih besar dalam penilaian kinerja. Begitu pula dengan kriteria lama kerja dengan nilai koefisien terbesar menandakan pengaruh yang besar terhadap peluang seorang pegawai mendapatkan promosi. Hal ini sejalan dengan aturan yang berlaku, dimana seorang pegawai mendapatkan kesempatan kenaikan golongan atau promosi setiap dua tahun sekali dengan hasil penilaian kinerja yang baik.

Hal ini berbanding terbalik dengan kriteria kepribadian (K1), sosial (K2), profesional (K4), dan keterlambatan yang menunjukkan tidak adanya pengaruh terhadap penilaian pegawai. Hal ini tentu saja menunjukkan perlu adanya perbaikan aspek-aspek penilaian pada kriteria-kriteria penilaian tersebut. Kepribadian yang baik, kemampuan bersosialisasi yang baik dan profesionalitas yang tinggi seorang pegawai tentu saja merupakan hal penting juga yang harus dimiliki seorang pegawai. Demikian pula dengan keterlambatan pegawai. Dalam pelaksanaannya ternyata banyaknya keterlambatan yang dilakukan seorang pegawai tidak berpengaruh terhadap penilaian kinerja. Hal ini tentu saja akan mengakibatkan ketidakadilan bagi pegawai, terutama para pegawai yang tepat waktu hadir di tempat kerja mendapatkan nilai yang sama dengan pegawai yang sering terlambat.

IV. SIMPULAN

Dari penelitian yang telah dilakukan diperoleh kesimpulan bahwa dimungkinkan untuk membentuk suatu model *decision tree* dan regresi logistik dari data yang ada. Dari seluruh penjabaran penelitian terlihat bagaimana proses klasifikasi data penilaian pegawai di Yayasan Prasama Bhakti dengan menggunakan kedua model tersebut untuk memprediksi kinerja pegawai sebagai rekomendasi promosi atau kenaikan golongan pegawai. Data induk, hasil penilaian DP3, kehadiran dan keterlambatan pegawai yang digunakan dalam penelitian kali ini dirasa cukup untuk memprediksi kinerja pegawai dengan membentuk *decision tree* dan regresi logistik. Berdasarkan nilai akurasi yang diperoleh dari masing-masing model, dapat diputuskan bahwa model *decision tree* menjadi model yang paling tepat dalam memprediksi kinerja pegawai dengan kriteria yang berlaku di Yayasan Prasama Bhakti. Nilai akurasi *decision tree* menunjukkan nilai yang lebih baik dari regresi logistik walaupun dengan perbedaan yang tipis. Hal ini membuktikan bahwa dengan menggunakan *decision tree* hasil prediksi lebih banyak yang sesuai dengan data sebenarnya daripada menggunakan regresi logistik.

Terlepas dari data awal yang sulit untuk disiapkan, dari seluruh data tersebut dapat dibentuk suatu model *decision tree* dengan nilai akurasi yang baik, begitu pula dengan model regresi logistik dengan nilai akurasi yang baik. Hal ini terbukti dengan model *decision tree* maupun regresi logistik yang berhasil memprediksi sebagian besar data tes yang ada sama dengan data yang sebenarnya.

Dari hasil perhitungan regresi logistik juga diketahui bahwa tidak seluruh atribut atau variabel penilaian pegawai berpengaruh terhadap keputusan akhir yayasan dalam memberikan kesempatan promosi atau kenaikan golongan seorang pegawai. Lama kerja seorang pegawai menjadi variabel yang paling berpengaruh. Hal ini sejalan dengan apa yang diterapkan oleh pihak yayasan, dimana seorang pegawai dengan lama kerja kelipatan dua tahun diberikan kesempatan untuk mendapatkan promosi atau kenaikan golongan. Sedangkan penilaian DP3 yang dilakukan oleh yayasan terhadap pegawainya hanya memberikan sedikit pengaruh pada kesempatan itu. Hal ini dapat menjadi catatan khusus bagi pihak Yayasan Prasama Bhakti untuk merancang sebuah penilaian dengan kriteria yang adil untuk pegawai-pegawainya.

Daftar Pustaka

- [1] D. Wibisono, *Manajemen Kinerja Korporasi dan Organisasi Panduan Penyusunan Indikator*, Jakarta: Erlangga, 2011.
- [2] T. McLean, "The measurement and management of human performance in seventeenth century English farming: The case of Henry Best," *Elsevier Accounting Forum* 33, pp. 62 - 73, 2009.
- [3] B. A. Groen, M. J. Wouters, & C. P. Wilderom, "Why do employees take more initiatives to improve their performance after co-developing performance measures? A field study," *Elsevier Management Accounting Research*, vol. 23, pp. 120 - 141, 2012.
- [4] K. Platts & M. Sobótka, "When the uncountable counts: An alternative to monitoring employee performance," *Business Horizons*, vol. 53, p. 349 - 357, 2010.
- [5] N. Breffni, "Customer perceived control and the moderating effect of restaurant type on evaluations of restaurant employee performance," *International Journal of Hospitality Management*, vol. 27, p. 23-29, 2008.
- [6] D. L. Deadrick & D. G. Gardner, "Maximal and typical measures of job performance: An analysis of performance variability over time," *Human Resource Management Review*, vol. 18, p. 133-145, 2008.
- [7] M. J. Aitkenhead, "A Co-evolving Decision Tree Classification Method," *Expert Systems with Applications*, vol. 34, pp. 18 - 25, 2008.
- [8] B. Aviad & R. Gelbard, "Classification by clustering decision tree-like classifier based on adjusted clusters," *Expert Syst. Appl.*, vol. 38, p. 8220-8228, 2011.
- [9] R. D. Yuniarti & R. Goejantoro, "Perbandingan Metode Klasifikasi Regresi Logistik Dengan Jaringan Saraf Tiruan," *EKSPONENSIAL*, vol. 4, no. 1, pp. 17-24, 2013.
- [10] I. Choina, R. Aulia, & A. Zakir, "Penerapan Algoritma ID3 Untuk Menyeleksi Pegawai Kontrak Di Kantor Pengadilan Kota Langsa," *CESS (Journal of Computer Engineering System and Science)*, vol. 5, no. 1, pp. 47-52, 2020.
- [11] N. Kristiyani, A. R. Tanaamah, & C. Fibriani, "Sistem Pendukung Keputusan dengan Menggunakan Algoritma Iterative Dichotomizer Three," *Jurnal Teknologi Informasi-Aiti*, vol. 8, no. 1, pp. 1-20, 2011.
- [12] R. G. Santosa & A. R. Chrismanto, "Perbandingan Akurasi Model Regresi Logistik untuk Prediksi Kategori IP Mahasiswa Jalur Prestasi dengan Non Jalur Prestasi," *Jurnal Teknik dan Ilmu Komputer*, vol. 7, no. 25, pp. 107-121, 2018.
- [13] N. H. Prabowo & Irhamah, "Klasifikasi Kinerja Karyawan Terhadap Penjualan Produk Menggunakan Metode Regresi Logistik dan Naïve Bayes," *INFERENCE*, vol. 1, no. 2, pp. 83-87, 2018.
- [14] S. Adelman & L. T. Moss, *Data Warehouse Project Management*, Canada: Addison-Wesley, 2000.
- [15] Nagitec, "Apa Itu Data Mining?," Nagitec, 2020. [Online]. Available: <https://nagitec.com/apa-itu-data-mining/>. [Accessed 20 Januari 2022].
- [16] M. Ayub, "Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer," *Jurnal Sistem Informasi*, vol. 2, no. 1, pp. 21 - 30, 2007.
- [17] B. D. M. Achmad & F. Slamet, "Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree," *Jurnal IPTEK*, vol. 16, no. 1, pp. 17 - 23, 2012.
- [18] Wahyudin, "Metode Iterative Dichotomizer 3 (ID3) Untuk Penyeleksian Penerimaan Mahasiswa Baru," *Jurnal Pendidikan Teknologi Informasi dan Komunikasi*, vol. 2, no. 2, pp. 5 - 15, 2009.
- [19] A. K. Nugroho & D. Iskandar, "Algoritma Iterative Dichotomizer 3 (ID3) Pengambilan Keputusan," *Dinamika Rekayasa*, vol. 11, no. 2, pp. 43-47, 2015.
- [20] Fractal, *Comparative Analysis of Classification Techniques*, A Fractal White Paper, 2003.
- [21] O. D. S. MALANG, "Uji Regresi Binary Logistik," Olah Data Statistik Malang, Januari 2018. [Online]. Available: <https://olahdatadimalang.blogspot.com/2018/01/uji-regresi-binary-logistik-maksimal-5.html>. [Accessed 20 Januari 2022].
- [22] A. Agresi, *Categorical Data Analysis Third Edition*, New Jersey: John Wiley & Sons, 2013.
- [23] M. Aryuni, "Tahap-Tahap Data Mining," Binus, 15 Desember 2016. [Online]. Available: <https://sis.binus.ac.id/2016/12/15/tahap-tahap-data-mining/>. [Accessed 20 Januari 2022].
- [24] Chairani & Syahputri R., "Penerapan Algoritma Iterative Dichotomizer 3 (ID3) untuk Penetapan Kelayakan Perubahan Status Kerja Karyawan," in *Seminar Nasional Ilmu Komputer*, Semarang, 2014.

LAMPIRAN
 Model Decision Tree Penilaian Kinerja Pegawai Yayasan Prasama Bhakti

