

Klasifikasi Perpanjangan Kontrak Karyawan pada PT Milpo Menggunakan Algoritma Naive bayes

¹Indah Kurniawati, ²Asep Muhidin, ³Fibi Eko Putra
^{1,2,3}Universitas Pelita Bangsa
Bekasi, Indonesia

¹indahhkurnia97@gmail.com, ²asep.muhammad@pelitabangsa.ac.id,
³fibi@pelitabangsa.ac.id

*Penulis Korespondensi

Diajukan : 05/01/2024

Diterima : 08/01/2024

Dipublikasi : 10/01/2024

ABSTRAK

Perkembangan ilmu pengetahuan dan teknologi yang ditandai dengan kemajuan di bidang teknologi komunikasi dan informasi saat ini telah berkembang begitu pesat. Peranan teknologi informasi sendiri pada aktivitas manusia pada saat ini begitu besar telah menjadi fasilitas utama bagi kegiatan berbagai sektor kehidupan dimana memberikan andil besar terhadap perubahan-perubahan yang semakin berkembang. Dalam setiap perusahaan untuk meningkatkan kinerja dan produktivitas secara maksimal, maka perusahaan dapat memberlakukan penilaian terhadap kinerja tenaga kerja. Perpanjangan kontrak karyawan merupakan agenda yang dilaksanakan oleh perusahaan dimana tenaga kerja yang memiliki kinerja yang baik bagi perusahaan berhak mendapat kontrak untuk masa waktu kedepan. Perusahaan dalam hal memberikan kesempatan perpanjangan masa kerja tentunya harus melakukan penilaian terhadap kinerja seluruh karyawannya terlebih dahulu, menilai karyawan bukan hal yang mudah bila jumlah karyawan terlalu banyak. Pada perusahaan kegiatan penilaian kinerja karyawan sulit dilaksanakan karena frekuensi tatap muka antara pihak manager dan karyawan sangat minim serta sulit menentukan nilai akurasi tingkat prediksi yang mempengaruhi perpanjangan kontrak karyawan. Pada penelitian ini diusulkan penggunaan teknik data mining untuk klasifikasi perpanjangan kontrak karyawan. Teknik data mining untuk memprediksi berdasarkan data karyawan PT.Milpo Indonesia bagian produksi. Metode data mining yang digunakan untuk memprediksi adalah klasifikasi, yaitu *Naive Bayes*. Dari penelitian ini *Naive Bayes* berhasil memprediksi 45 karyawan yang diperpanjang kontrak dan 41 karyawan yang diputus kontrak dari 86 karyawan dengan nilai accuracy sebesar 98.75%, nilai Precision sebesar 98.00% dan nilai Recall sebesar 100%, sehingga metode naive bayes merupakan metode yang cukup baik dalam penelitian ini.

Kata Kunci: Data Mining, Klasifikasi, *Naive Bayes*, Perpanjangan Kontrak.

I. PENDAHULUAN

Dengan kemajuan dalam teknologi komunikasi dan informasi saat ini, perkembangan ilmu pengetahuan dan teknologi telah berkembang dengan sangat cepat. Pada saat ini, peran teknologi informasi begitu besar dalam aktivitas manusia, sehingga telah menjadi sarana penting untuk berbagai bagian kehidupan yang menghasilkan perubahan yang semakin besar. Berbagai kelompok berusaha untuk menguasai dan mengembangkan teknologi melalui kegiatan penelitian dan pengembangan. Selain itu, data dikumpulkan untuk membantu kemajuan teknologi. Perusahaan-perusahaan di antaranya terus memperbarui teknologi internal mereka untuk tetap mengikuti perkembangan zaman. (Heliyanti Susana, 2022) Dalam setiap perusahaan Untuk meningkatkan produktivitas dan kinerja, bisnis dapat melakukan penilaian kinerja tenaga kerja dan menerapkan sistem tenaga kerja kontrak. Jika perusahaan ingin memberikan kesempatan

perpanjangan waktu kerja kepada karyawannya, mereka harus menilai kinerja semua karyawannya terlebih dahulu. Menilai kinerja karyawan di perusahaan yang memiliki terlalu banyak karyawan akan sulit. Hal ini akan menjadi tantangan bagi tim penilai untuk mengolah data dan mencapai hasil terbaik, terutama karena ada banyak kriteria penilaian yang akan membuat semua data karyawan diperiksa dengan teliti dan hati-hati. Keputusan manajer sangat penting bagi perusahaan untuk dampak selanjutnya. Ini karena perusahaan tidak memiliki banyak interaksi tatap muka antara manajer dan karyawan, dan sulit untuk menentukan tingkat keakuratan prediksi yang mempengaruhi durasi kontrak karyawan. Perpanjangan tenaga kerja kontrak adalah rencana yang dibuat oleh perusahaan untuk memberikan pekerjaan kepada tenaga kerja yang memiliki kinerja baik untuk jangka waktu yang lebih lama. Untuk memastikan bahwa perusahaan juga mendapatkan sumber daya manusia terbaik. (Hadikristanto & Fitri, 2018) Metode data mining yang dikenal sebagai klasifikasi membuat model sampel data yang belum terklasifikasi yang dapat digunakan untuk mengklasifikasi sampel data baru ke dalam kelas yang sebanding. Klasifikasi termasuk dalam pengajaran *supervised* karena menggunakan sekumpulan data untuk dianalisis terlebih dahulu, dan kemudian pola dari hasil analisis tersebut digunakan untuk mengklasifikasikan data uji. Proses klasifikasi data terdiri dari dua bagian: pembelajaran data dan klasifikasi. Algoritma klasifikasi digunakan untuk menganalisis pembelajaran data, dan pengujian data digunakan untuk memastikan bahwa peraturan klasifikasi yang digunakan cukup akurat. Teknik klasifikasi dibagi menjadi lima kategori berdasarkan perbedaan konsep matematika: statistik, jarak, pohon keputusan, jaringan syaraf, dan *rule*. Ada banyak algoritma untuk kategori ini, tetapi yang paling umum adalah naive bayes. Performa pengklasifikasi biasanya diukur dengan ketepatan atau tingkat galat. (Rachman & Handayani, 2021) *Naive Bayes* (NB) adalah algoritma klasifikasi probabilitas sederhana yang berdasarkan pada teorema *Bayes*, asumsi bebas yang kuat (*naive*), dan model fitur independen. *Naive Bayes* juga merupakan algoritma klasifikasi yang utama pada data mining dan banyak diterapkan dalam masalah klasifikasi di dunia nyata karena memiliki performa klasifikasi yang tinggi. Menghitung peluang untuk suatu hipotesis, menghitung peluang dari suatu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal. Berdasarkan uraian di atas maka penulis tertarik untuk melakukan penelitian dengan judul “Klasifikasi Perpanjangan Kontrak Karyawan Pada PT Milpo Indonesia Menggunakan Algoritma Naive Bayes”.

II. STUDI LITERATUR

Penelitian Terdahulu

Penelitian ini menerapkan Data Mining dengan menggunakan Algoritma *Naive Bayes* untuk klasifikasi perpanjangan kontrak kerja karyawan. Tujuannya adalah membantu menentukan apakah kontrak kerja karyawan sebaiknya diperpanjang atau tidak, dengan menggunakan delapan variabel seperti absensi, disiplin, tanggung jawab, kebersihan, sopan santun, umur, kerja sama tim, dan inisiatif kerja. Algoritma *Naive Bayes* menghasilkan tingkat akurasi sebesar 90% dalam mengklasifikasikan perpanjangan kontrak kerja karyawan berdasarkan data karyawan yang digunakan dalam penelitian (Titin Winarti, 2020) Penelitian ini menggunakan metode *Naive Bayes* untuk mengklasifikasikan ketepatan waktu kelulusan mahasiswa Fakultas Ekonomi Universitas Pamulang pada tahun akademik 2018/2019. Atribut yang digunakan dalam klasifikasi melibatkan ketepatan waktu kelulusan, daerah asal, jenis kelamin, profesi orang tua, jenis sekolah, program studi, dan predikat nilai rata-rata kumulatif (IPK). Hasil penelitian menunjukkan bahwa sebanyak 61,9% mahasiswa S1 Fakultas Ekonomi Universitas Pamulang lulus tepat waktu, dengan tingkat akurasi klasifikasi mencapai 69,33% (Tias Mugi Rahayu, 2021). Penelitian ini fokus pada prediksi kelayakan calon penerima bantuan bidikmisi pengganti dari Universitas Gadjah Mada dengan menggunakan Algoritma *Naive Bayes*. Menggunakan pemodelan data mining, penelitian ini sistematis mempelajari langkah-langkah untuk memprediksi kelayakan calon penerima bidikmisi. Data primer dari pelamar bantuan bidikmisi UGM digunakan, dengan kriteria seperti penghasilan orang tua, jumlah tanggungan orang tua, IPK, dan status penerimaan beasiswa. Pengujian model menunjukkan akurasi tinggi sebesar 93,4%, menandakan bahwa Algoritma *Naive Bayes* dapat diandalkan sebagai alternatif pengambilan keputusan untuk memprediksi kelayakan calon penerima bantuan bidikmisi

pengganti (AISYAH, 2021)

Naive Bayes

Algoritma *Naive Bayes* merupakan salah satu algoritma dari algoritma klasifikasi statistik. Metode klasifikasi menggunakan probabilitas dan metode statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu prediksi peluang masa depan berdasarkan pengalaman masa lalu, oleh karena itu dinamai teorema *Bayes*. Ciri utama pengklasifikasi *Naive Bayes* ini adalah asumsi (naif) yang sangat kuat tentang independensi setiap kondisi/peristiwa. (Wibowo, 2022)

Rumus dari theorem *bayes* sendiri ialah :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan :

X : Data dengan kelas yang belum diketahui

H : Hipotesis data x merupakan suatu kelas spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)

P(H) : Probabilitas hipotesis H (*prior probability*)

P(X|H) : Probabilitas X berdasar kondisi pada hipotesis

H P(X) : Probabilitas X

Cross Validation

Pada dasarnya, *cross-validation* (CV) adalah metode statistik yang dapat digunakan untuk mengevaluasi kinerja suatu model atau algoritma di mana data dipisahkan menjadi dua subset, yaitu data regresi, pelatihan, dan data validasi/penilaian. dua bagian dengan satu bagian digunakan sebagai data pelatihan dan bagian lainnya digunakan sebagai data uji. (Nugroho, 2020) Beberapa penelitian membagi data menjadi 10 bagian, 90% digunakan sebagai pelatihan dan 10 sisanya digunakan sebagai tes. Proses ini diulangi hingga 10 kali hingga semua record data menjadi bagian dari data uji. Proses ini juga dikenal sebagai validasi silang 10 kali lipat. Validasi silang 10 kali lipat banyak digunakan oleh para peneliti karena telah terbukti menghasilkan kinerja algoritma yang lebih stabil. *K-fold* merupakan salah satu metode *cross validation*. (Nugroho, 2020)

Evaluasi Model

Confusion matrix, yang juga disebut sebagai *error matrix*, memberikan data tentang perbandingan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya. Tabel *matriks* yang menggambarkan kinerja model klasifikasi pada set data uji yang nilai sebenarnya diketahui dikenal sebagai *matriks confusion*. Salah satu fungsi matriks *confusion* adalah untuk mengukur tingkat ketepatan, ketepatan, dan nilai *recall* dari model algoritma yang dievaluasi. Nilai ketepatan adalah tingkat ketepatan presentase antara nilai prediksi dan nilai sebenarnya, dan nilai ketepatan adalah nilai akurasi dengan kelas yang diprediksi. Selain itu, nilai *recall* adalah presentase nilai kinerja keberhasilan algoritma yang digunakan *matriks*, yang menunjukkan kinerja model klasifikasi pada serangkaian data uji yang sebenarnya di ketahui. (Muhidin, 2023)

Rapid Miner

RapidMiner adalah salah satu alat data mining yang digunakan, dengan lebih dari 500 operator untuk *input*, *output*, *preprocessing*, dan visualisasi data. Ini juga merupakan *software* yang berdiri sendiri untuk analisis data dan mesin data mining yang dapat diintegrasikan pada produknya sendiri. (Reyvan Maulid, 2023). Pada tahun 2010-2011, *KDnuggets*, sebuah portal data mining, menilai *RapidMiner* sebagai *software* data mining terbaik. *RapidMiner* menggunakan GUI (*Graphic User Interface*) untuk membuat pipeline analitis. GUI ini menghasilkan file XML (*Extensible Markup Language*) yang menjelaskan proses analitis yang diinginkan pengguna untuk diterapkan ke data. *RapidMiner* kemudian membaca file ini untuk memulai analisis secara otomatis. (Widiyaningrum, 2020)

III. METODE

Metode Pengumpulan Data

Pada langkah berikutnya, setelah melakukan pengamatan dan observasi, peneliti sudah menentukan untuk menggunakan metode algoritma *naive Bayes*. Dalam penelitian ini akan dilakukan beberapa langkah atau tahapan penelitian.



Gambar 3.1 Tahapan penelitian

Dari gambar diatas dapat diuraikan tahapannya sebagai berikut :

Pengumpulan Data

Dalam penelitian ini, dua jenis data yang digunakan adalah kualitatif (kata-kata) dan kuantitatif (angka). Data kuantitatif mencakup berbagai aspek seperti kehadiran, tanggung jawab, pengetahuan, loyalitas, ketelitian, kerjasama, kepercayaan, kreativitas, ekspresif, keamanan, dan inisiatif. Selain itu, data kuantitatif juga mencakup atribut seperti nama, jenis kelamin, departemen, bagian, grup, kontrak, dan jangka waktu yang diberikan kepada karyawan. Tabel yang disediakan menunjukkan rincian parameter dari kedua jenis data tersebut. Meskipun data kuantitatif tidak berpengaruh saat pengujian, namun perlu dibersihkan sebelum analisis lebih lanjut

Pengolahan Data Awal

Data karyawan yang telah dikumpulkan dibersihkan, yang berarti menghilangkan rekaman yang tidak jelas atau tidak lengkap serta rekaman berulang. Informasi yang tidak diperlukan, seperti NIK, nama karyawan, jenis kelamin, departemen, bagian, grup, kontrak, dan jangka waktu, juga dihilangkan. Hal ini dilakukan karena informasi ini tidak akan berdampak pada proses pengolahan data berikutnya. Tabel atribut dan kategorinya dapat ditemukan pada tabel 3.4.

Dalam persiapan untuk klasifikasi, data yang telah disiapkan dibagi menjadi dua bagian, yaitu data training (90%) dan data *testing* (10%). Pembagian ini dilakukan melalui teknik sampling random sistematis, di mana perandoman dilakukan hanya sekali saat menentukan komponen pertama dari sampel. Interval sampel, biasanya disimbolkan dengan huruf k , digunakan untuk menentukan unsur sampel berikutnya dalam kerangka sampel, menciptakan pemilihan yang sistematis untuk keperluan pengujian. Interval sampel atau juga disebut sampling rasio diperoleh dengan cara membagi ukuran populasi dengan ukuran sampel yang dikehendaki (N/n). Contoh perhitungan untuk mengambil data testing adalah sebagai berikut:

$$\text{Jumlah data testing} = 10\% \times 86 = 9$$

$$\text{Jumlah populasi (N)} = 86$$

$$\text{Jumlah sampel (n)} = 9$$

Interval Sampling (k) = $N/n = 86/9 = 9$

Unsur pertama yang diambil untuk data testing (s) = 1

Unsur kedua = s + k

Unsur ketiga = s + 2k

Unsur keempat = s + 3k, dan seterusnya hingga unsur ke-n

Dari hasil diatas diperoleh data testing sebanyak 9 data karyawan, maka sisanya dijadikan data training sebanyak 77 data karyawan.

Metode yang di usulkan

Dalam penelitian ini, metode algoritma naive Bayes akan digunakan untuk melakukan analisis. Data akan dihitung dengan metode ini, dan kemudian akan dicari hasil akurasi algoritma. Saat ini, beberapa langkah pengujian data akan dilakukan, pengolahan data dengan *naive bayes*, hasil algoritma *naive bayes*, Melihat Hasil Akurasi Metode Algoritma *Naive Bayes* dengan *RapidMiner*.

IV. HASIL DAN PEMBAHASAN

Hasil

1. Evaluasi dan Validasi Hasil Algoritma Naive Bayes

Pada langkah awal dari proses algoritma naive bayes ini adalah siapkan terlebih dahulu data yang akan diproses.

Tabel 4. 1 Data Keseluruhan (Nama, Atribut yang bernilai angka dan hasil)

NAMA KARYAWAN	KEHADIRAN	TANGGUNG JAWAB	PENGETAHUAN	LOYALITAS	KETELITIAN	KERJASAMA	KEPERCAYAAN	KREATIVITAS	EKSPRESIF	SAFETY	INISITIF	HASIL KEPUTUSAN
ABAY	4	3	3	3	4	2	2	3	2	3	3	DIPERPANJANG KONTRAK
ACEP	4	2	2	3	3	4	2	3	1	2	3	DIPERPANJANG KONTRAK
ADE HIDAYAT	1	1	2	2	1	2	2	2	1	2	1	DIPUTUS KONTRAK
ADUL	1	1	1	1	1	1	2	1	2	2	2	DIPUTUS KONTRAK
AGUNG SETYADI	3	3	1	1	1	1	1	1	1	2	1	DIPUTUS KONTRAK
AHMAD FAUZI	4	4	3	3	3	4	4	4	4	3	4	DIPERPANJANG KONTRAK
ALI SUHARDI	4	3	3	3	3	2	2	3	2	3	3	DIPERPANJANG KONTRAK
ALIF KOMARUDIN	3	4	4	3	2	4	4	4	4	3	2	DIPERPANJANG KONTRAK
ANDHIKA SEPTIANO	4	4	3	4	3	2	2	3	2	3	3	DIPERPANJANG KONTRAK
ANDRI HIDAYAT	4	4	4	4	3	3	3	2	2	3	4	DIPERPANJANG KONTRAK
ANGGITA LESTARI	2	4	1	1	2	1	1	1	1	1	2	DIPUTUS KONTRAK
ANIN	2	3	1	2	1	1	1	1	1	1	1	DIPUTUS KONTRAK
AGUSTIN	4	3	3	3	3	2	2	3	2	3	3	DIPERPANJANG KONTRAK
ANIS SAPUTRA	2	2	2	1	2	1	1	1	1	2	1	DIPUTUS KONTRAK
APRILIAN ZAHIRA	3	4	3	4	2	4	4	2	2	3	4	DIPERPANJANG KONTRAK
ARDIANSYAH	4	4	3	3	4	2	2	4	2	4	3	DIPERPANJANG KONTRAK
ARI PANJI	4	4	4	3	4	4	2	4	2	4	3	DIPERPANJANG KONTRAK
ARIF LUKMAN	4	3	4	3	3	4	2	4	2	4	4	DIPERPANJANG KONTRAK
ARUN	4	4	4	3	4	2	3	4	4	4	3	DIPERPANJANG KONTRAK
ASTRI LESTARI	3	1	2	1	1	1	3	2	1	1	1	DIPUTUS KONTRAK

Data diatas adalah data keseluruhan, untuk melakukan pengolahan data di *RapidMiner* data yang digunakan adalah data *training*, dari seluruh populasi, data di pisah menjadi 90% data training yang di pakai untuk membuat model atau rule dan 10% sebagai data testing yang dipakai untuk sample testing.

Tabel 4. 2 Data Training (90% Dari Data Keseluruhan)

KEHADIRAN	TANGGUNG JAWAB	PENGETAHUAN	LOYALITAS	KETELITIHAN	KERJASAMA	KEPERCAYAAN	KREATIVITAS	EKSPRESIF	SAFETY	INISIATIF	HASIL KEPUTUSAN
4	3	3	3	4	2	2	3	2	3	3	DIPERPANJANG KONTRAK
4	2	2	3	3	4	2	3	1	2	3	DIPERPANJANG KONTRAK
1	1	2	2	1	2	2	2	1	2	1	DIPUTUS KONTRAK
1	1	1	1	1	1	2	1	2	2	2	DIPUTUS KONTRAK
3	3	1	1	1	1	1	1	1	2	1	DIPUTUS KONTRAK
4	4	3	3	3	4	4	4	4	3	4	DIPERPANJANG KONTRAK
4	3	3	3	3	2	2	3	2	3	3	DIPERPANJANG KONTRAK
3	4	4	3	2	4	4	4	4	3	2	DIPERPANJANG KONTRAK
4	4	3	4	3	2	2	3	2	3	3	DIPERPANJANG KONTRAK
4	4	4	4	3	3	3	2	2	3	4	DIPERPANJANG KONTRAK
2	4	1	1	2	1	1	1	1	1	2	DIPUTUS KONTRAK
2	3	1	2	1	1	1	1	1	1	1	DIPUTUS KONTRAK
4	3	3	3	3	2	2	3	2	3	3	DIPERPANJANG KONTRAK
2	2	2	1	2	1	1	1	1	2	1	DIPUTUS KONTRAK
3	4	3	4	2	4	4	2	2	3	4	DIPERPANJANG KONTRAK
4	4	3	3	4	2	2	4	2	4	3	DIPERPANJANG KONTRAK
4	4	4	3	4	4	2	4	2	4	3	DIPERPANJANG KONTRAK
4	3	4	3	3	4	2	4	2	4	4	DIPERPANJANG KONTRAK
4	4	4	3	4	2	3	4	4	4	3	DIPERPANJANG KONTRAK
3	1	2	1	1	1	3	2	1	1	1	DIPUTUS KONTRAK
4	4	3	4	3	3	3	3	3	3	3	DIPERPANJANG KONTRAK
2	2	2	1	2	1	2	2	2	1	4	DIPERPANJANG KONTRAK
4	3	2	3	1	1	2	1	2	1	1	DIPUTUS KONTRAK
3	2	1	1	2	2	2	3	3	2	2	DIPUTUS KONTRAK
4	2	2	1	1	1	2	2	1	1	2	DIPUTUS KONTRAK
4	4	3	3	2	3	4	3	2	3	4	DIPERPANJANG KONTRAK
4	3	3	3	3	2	2	3	2	3	3	DIPERPANJANG KONTRAK
4	4	4	4	3	4	3	4	3	4	3	DIPERPANJANG KONTRAK
3	3	1	1	1	2	2	2	2	1	2	DIPUTUS KONTRAK
4	3	4	3	4	2	4	3	3	4	4	DIPERPANJANG KONTRAK

Data diatas adalah data *Training*, data *Training* berjumlah 90% dari total data, jumlah



data Training sebanyak 77 karyawan. Data training dipilih secara acak dan atribut nama dihilangkan karena atribut nama tidak berpengaruh saat memproses data di *RapidMiner*. Atribut yang digunakan hanya kehadiran, tanggung jawab, pengetahuan, loyalitas, keseriusan, kerjasama, kepercayaan, kreativitas, ekspresif, keamanan, inisiatif dan komitmen, konsekuensi keputusan (perpanjangan atau pemutusan kontrak).

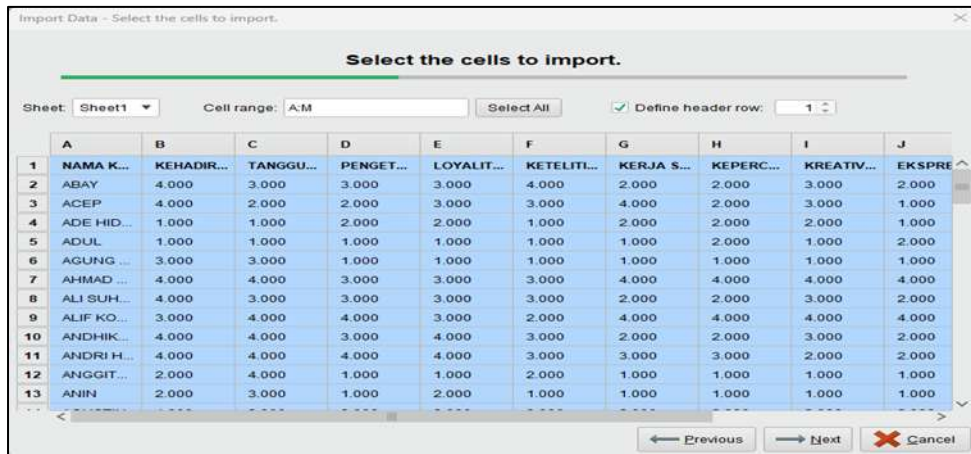
Tabel 4. 3 Data Testing (10% Dari Data Keseluruhan)

4	3	3	3	4	2	2	3	2	3	3	DIPERPANJANG KONTRAK
4	2	2	3	3	4	2	3	1	2	3	DIPERPANJANG KONTRAK
1	1	2	2	1	2	2	2	1	2	1	DIPUTUS KONTRAK
1	1	1	1	1	1	2	1	2	2	2	DIPUTUS KONTRAK
3	3	1	1	1	1	1	1	1	2	1	DIPUTUS KONTRAK
4	4	3	3	3	4	4	4	4	3	4	DIPERPANJANG KONTRAK
4	3	3	3	3	2	2	3	2	3	3	DIPERPANJANG KONTRAK
3	4	4	3	2	4	4	4	4	3	2	DIPERPANJANG KONTRAK
4	4	3	4	3	2	2	3	2	3	3	DIPERPANJANG KONTRAK

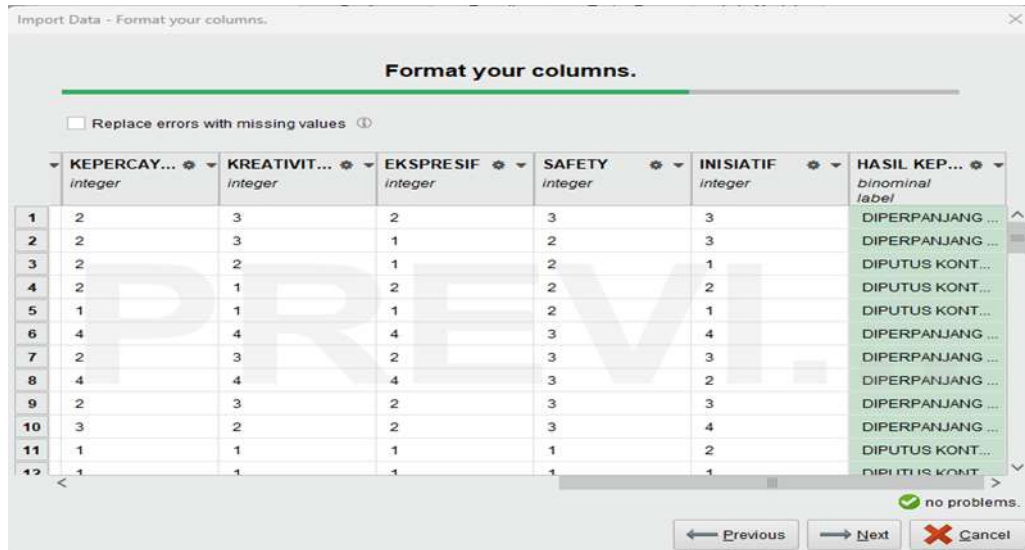
Data diatas merupakan data tes, data tes berjumlah 10 item data, jumlah data tes seluruhnya 9 orang pegawai. Data uji diacak dan atribut nama dihilangkan karena atribut nama tidak berpengaruh saat proses data di *RapidMiner*. Atribut yang digunakan hanya kehadiran, tanggung jawab, pengetahuan, loyalitas, keseriusan, kerjasama, kepercayaan, kreativitas, ekspresif, keamanan, inisiatif dan komitmen, konsekuensi keputusan (perpanjangan atau pemutusan kontrak).

2. Pengolahan Data *Training*

Langkah awal mengolah data *training* adalah, *import* data *training* ke dalam *tools RapidMiner*.

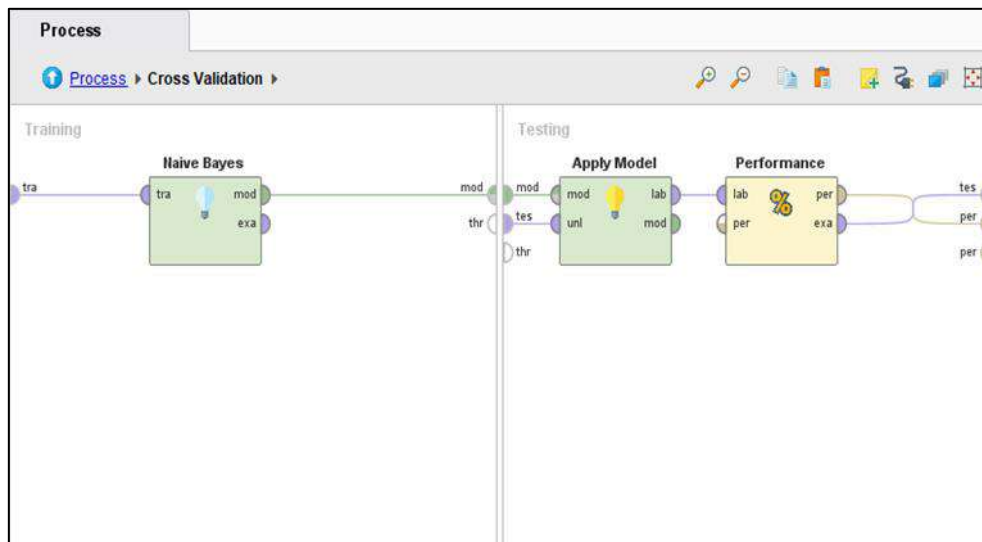


Gambar 4. 1 Pemilihan Data *Training*



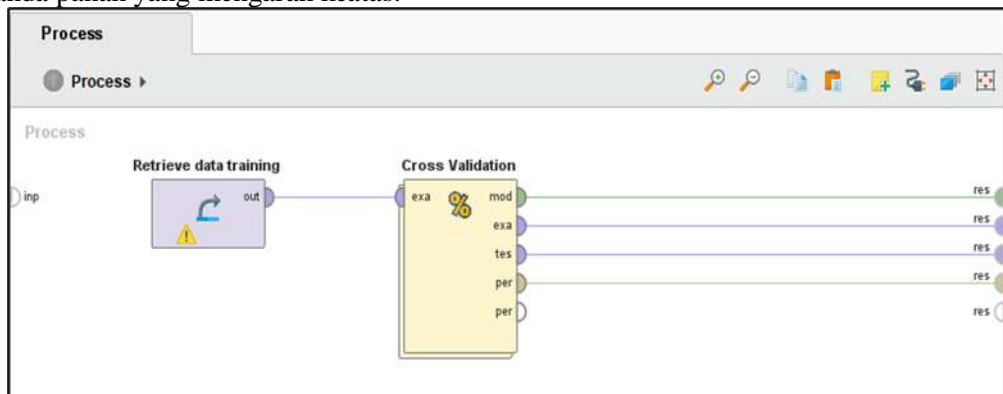
Gambar 4. 2 Pemilihan Hasil Keputusan sebagai Label

Pilih Hasil Keputusan sebagai binominal dan label, sedangkan yang lainnya sebagai *integer*. Setelah itu tekan *Next* lalu *Finish*.



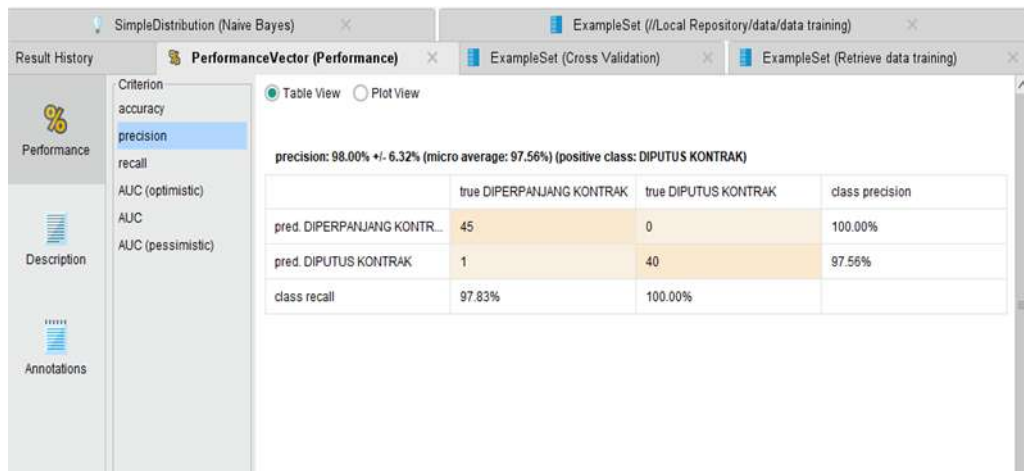
Gambar 4. 3 Main Proses Naive Bayes

Setelah muncul kotak *Naive Bayes*, *Apply Model* dan *Performance* Main Proses klik tanda panah yang mengarah keatas.



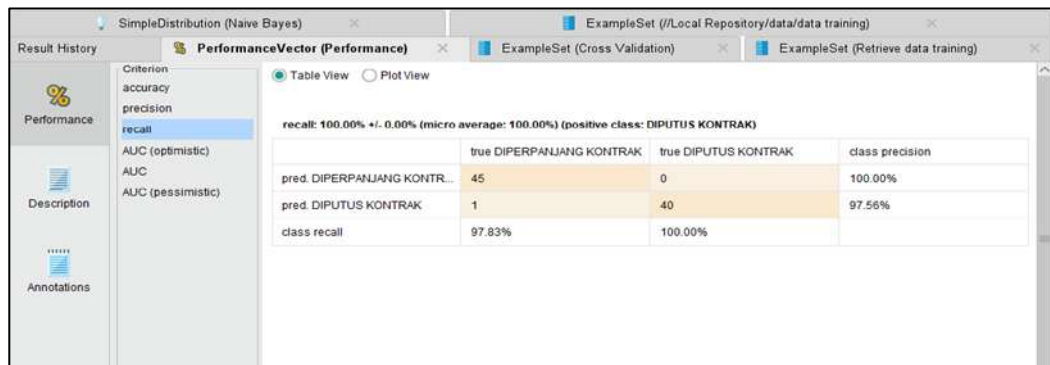
Gambar 4. 4 Main Proses Validation

Setelah muncul *Validaton* pada Main Proses geser *validation* ke kanan lalu sambungkan kabel *input* ke *Retrieve* satu kabel. *Retrieve* ke *Validation* satu kabel dan *Validation* ke res empat kabel. Lalu tekan running atau play (segitiga warna biru) di sebelah atas. Hasil pengukuran akurasi data yang diperoleh dari data latih mencapai 98,75% +/- 3,95% dan mikro 98,84%. Dari tabel ini, kita mengetahui bahwa perkiraan perpanjangan kontrak dengan true diperpanjang kontrak mencapai 45 karyawan dan true dihentikan 0 karyawan, debgan hasil nilai pasti 100, 00%. Sedangkan untuk memprediksi data pegawai yang diberhentikan, untuk kontrak riil yang diperpanjang sebanyak 1 pegawai dan untuk kontrak riil yang diberhentikan sebanyak 40 pegawai. Tingkat pemutusan kontrak karyawan sebesar 97,56%. Untuk tingkat *recall* data pegawai yang sangat tinggi terdapat class *recall true* dimana kontrak diputus mencapai 100,00%, sedangkan untuk class *recall true* dimana kontrak diperpanjang mencapai 97,00%,83%.



Gambar 4. 5 Pengukuran Data *Precision*

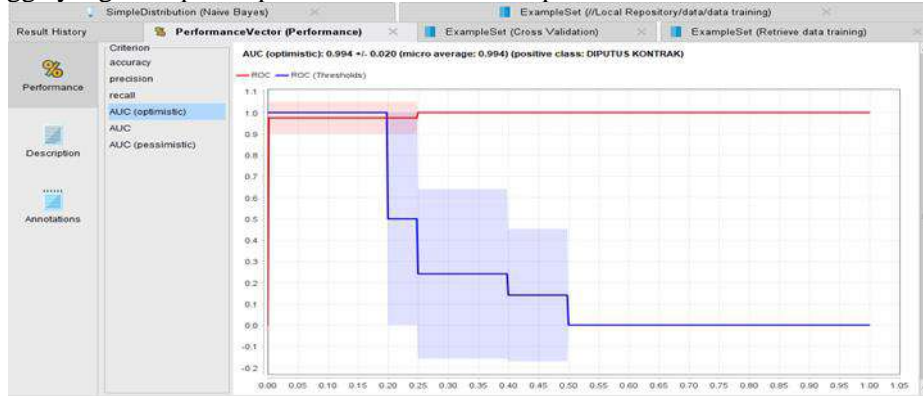
Hasil pengukuran akurasi menunjukkan nilai keseluruhan sebesar 98.00% dengan +/- 6.32% dan akurasi mikro mencapai 97.56%. Dari tabel hasil, diperoleh perkiraan bahwa dari 45 karyawan yang kontraknya diperpanjang, hasil prediksinya tepat, dan tidak ada karyawan yang seharusnya diakhiri kontrak. Sebagai prediksi pemutusan kontrak, terdapat 40 karyawan yang benar-benar berhenti dan 1 karyawan yang seharusnya tidak dihentikan. Tingkat akurasi pemutusan kontrak mencapai 97.56%. Recall data karyawan menunjukkan performa yang sangat baik, dengan recall 100.00% untuk karyawan yang seharusnya dihentikan dan 97.83% untuk karyawan yang seharusnya diperpanjang kontrak.



Gambar 4. 6 Pengukuran *Recall* Data

Hasil pengukuran *recall data* adalah 100.00% dengan +/-0.00% serta mikri 100.00%.

Jika diperhatikan untuk *recall data* prediksi adalah sama dengan *accuracy true* diputus kontrak dan *precision true* diputus kontrak. Yang membedakannya dengan yang lain adalah nilai mikro dari recall lebih tinggi yaitu sebesar 100.00%. Nilainya mengartikan bahwa untuk nilai atau hasil keputusan dalam true diperpanjang kontrak lebih kecil, sehingga yang merupakan *positive class* adalah “diputus kontrak”.



Gambar 4. 7 Pengukuran *Curva ROC AUC Optimistic*

Nilai dari *Curva ROC diukur* terlebih dahulu terhadap AUC dengan nilai sebesar 0.994 dengan +/-0.020 serta mikro 0.994 terhadap *positive class* diputus kontrak.



Gambar 4. 8 Pengukuran *Curva ROC AUC Normal*

Nilai *Curva ROC* terhadap AUC Normal dengan nilai sebesar 0.994 dengan +/- 0.020 serta mikro 0.994 terhadap *positive class* diputus kontrak.



Gambar 4. 9 Pengukuran *Curva ROC AUC Pessimistic*

V. KESIMPULAN

Berdasarkan pada hasil penelitian yang telah dilakukan maka dapat disimpulkan sebagai berikut:

1. Berdasarkan data bagian produksi PT.Milpo Indonesia, prediksi perpanjangan kontrak dengan metode *naive bayes* menunjukkan bahwa 45 karyawan diberi perpanjangan kontrak dan 41 karyawan di putus kontrak dari 86 karyawan.
2. Berdasarkan hasil perhitungan dengan *RapidMiner*, metode *Naive Bayes* memiliki nilai akurasi 98,75%, nilai ketepatan 98,00%, dan nilai recall 100.00%. Hasil ini menunjukkan bahwa metode *Naive Bayes* adalah metode yang cukup baik dalam penelitian tentang perpanjangan kontrak karyawan.

VI. REFERENSI

- (Titin Winarti, 2020) AISYAH, S., WAHYUNINGSIH, S., & AMIJAYA, F. (2021). Peramalan Jumlah Titik Panas Provinsi Kalimantan Timur Menggunakan Metode Radial Basis Function Neural Network. *Jambura Journal of Probability and Statistics*, 2(2), 64–74. <https://doi.org/10.34312/jjps.v2i2.10292>
- Hadikristanto, W., & Fitri, V. A. A. (2018). Penerapan Algoritma Naïve Bayes Berbasisparticle Swarm Optimization (Pso) Dalammenangani Kasus Kanker Payudara. *Jurnal Informatika*, 9(September), 111–120. <https://www.jurnal.pelitabangsa.ac.id/index.php/sigma/article/view/472%0Ahttps://www.jurnal.pelitabangsa.ac.id/index.php/sigma/article/download/472/286>
- Heliyanti Susana. (2022). Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, 4(1), 1–8. <https://doi.org/10.52005/jursistekni.v4i1.96>
- Muhidin, A., Danny, M., & Rilvani, E. (2023). Algoritme Multinomial Naïve Bayes Pada Aplikasi Chatbot Layanan Informasi Berbasis Teks. *Progresif: Jurnal Ilmiah Komputer*, 19(1), 71–80.
- Nugroho, K. S., Istiadi, I., & Marisa, F. (2020). Naive Bayes classifier optimization for text classification on e-government using particle swarm optimization. *Jurnal Teknologi Dan Sistem Komputer*, 8(1), 21–26. <https://doi.org/10.14710/jtsiskom.8.1.2020.21-26>
- Rachman, R., & Handayani, R. N. (2021). Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM. *Jurnal Informatika*, 8(2), 111–122. <https://doi.org/10.31294/ji.v8i2.10494>
- Titin Winarti, D. P. (2020). Penerapan Data Mining Menggunakan Algoritma Naive Bayes ntuk Klasifikasi Perpanjangan Kontrak Kerja Karyawan. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- Wibowo, M. F. S., Puspitasari, N. F., & Satya, B. (2022). Penerapan Data Mining Dan Algoritma Naïve Bayes Untuk Pemilihan Konsentrasi Mahasiswa Menggunakan Metode Klasifikasi. *Journal of Information System Management (JOISM)*, 3(2), 39–45. <https://doi.org/10.24076/joism.2022v3i2.680>
- Widiyaningrum, D. K., Septiani, R., Larasati, P., & Wibowo, A. (2020). Algoritma Naive Bayes untuk Menentukan Kelayakan Perpanjangan Kontrak pada PT. Gemilang. *Jurnal Ilmu Komputer Dan Bisnis*, 11(2), 2441–2445. <https://doi.org/10.47927/jikb.v11i2.5>