

Mata Kuliah : Data Warehouse dan Data Mining

Dosen Pengampu : Rita Sari, S.ST., M.Kom. / Rizqi Putri Nourma Budiarti, S.T., M.T.

Berdoalah terlebih dahulu sebelum mengerjakan soal ujian.

1. Jelaskan apa yang Anda ketahui tentang Data Mining? Jelaskan juga perbedaannya dengan Machine Learning.
2. Jelaskan bagaimana konsep dasar dari Classification dan sebutkan jenis beserta contohnya.
3. Jelaskan bagaimana konsep dasar dari Clustering dan sebutkan jenis beserta contohnya.
4. Jelaskan bagaimana proses Data Mining secara Supervised Learning maupun Unsupervised Learning, berikan contohnya.
5. Diketahui data training pada sebuah tabel:

NO	USIA	BATUK	SUHU BADAN	STATUS CORONA
1	TUA	KERING	TINGGI	POSITIF
2	TUA	KERING	TINGGI	NEGATIF
3	TUA	KERING	TINGGI	POSITIF
4	MUDA	KERING	TINGGI	NEGATIF
5	MUDA	KERING	RENDAH	POSITIF
6	MUDA	BASAH	RENDAH	NEGATIF
7	MUDA	BASAH	RENDAH	POSITIF
8	MUDA	BASAH	TINGGI	NEGATIF
9	TUA	BASAH	RENDAH	NEGATIF
10	TUA	KERING	RENDAH	POSITIF

Tabel ini hanya contoh, bukan sesuai dengan fakta kejadian.

Kerjakan dengan perhitungan manual algoritma Bayesian Classifier untuk memprediksikan hasil data testing berikut.

USIA	BATUK	SUHU BADAN	STATUS CORONA
TUA	BASAH	TINGGI	???

6. Terapkan perhitungan manual metode K-Mean Clustering pada data set berikut untuk 2 cluster. (single data set) Dataset { 4, 6, 10, 12, 3, 20, 40, 3x, 25 }

Keterangan: ganti x dengan 1 digit terakhir pada NIM Anda (tidak dikalikan)

7. Terapkan perhitungan manual metode K-Mean Clustering pada data set berikut untuk 2 cluster.

Sample No	X	Y
1	195	72
2	180	5(x)
3	178	60
4	189	68
5	192	72
6	198	77

Keterangan: ganti x dengan 1 digit terakhir pada NIM Anda (tidak dikalikan)

8. Temukan satu permasalahan yang dapat diselesaikan dengan proses data mining yang relevan dengan rincian berikut.

- 1. Jelaskan latar belakang dari permasalahan yang dipilih**
- 2. Jelaskan algoritma yang sesuai dengan permasalahan yang dipilih**
- 3. Buatlah dataset dari permasalahan yang dipilih**
- 4. Cari minimal 3 referensi yang mendukung dalam pemilihan permasalahan dan penyelesaian dengan algoritma yang sesuai**

Nama : Yogaraksa Amjad Hernawan

NIM : 3130020023

Jawaban UAS DWDM

1. Data Mining adalah suatu proses ekstraksi informasi atau pola yang penting atau menarik dari data yang penting dalam jumlah yang besar atau big data sehingga menjadi informasi yang sangat berharga.

Perbedaan pada Data Mining dengan Machine Learning adalah Jika data mining dirancang untuk mengekstrak aturan dari data dalam jumlah besar, sementara machine learning mengajarkan komputer cara belajar dan memahami parameter yang diberikan

2. Klasifikasi adalah sebuah teknik untuk mengklasifikasikan atau mengkategorikan beberapa item yang belum berlabel ke dalam sebuah set kelas diskrit.

Jenis - jenis klasifikasi yang sering digunakan dalam data science, yaitu:

- CONTENT-BASED CLASSIFICATION

Klasifikasi tipe satu ini berbasis konten, Pada tipe ini, pengelompokan data didasarkan pada jenis konten tiap file yang tersedia.

- USER-BASED CLASSIFICATION

Seperti namanya, classification ini menitikberatkan pada pengguna, tepatnya pemahaman pengguna terhadap pembuatan, penyuntingan, peninjauan, atau penyebaran data. Dari situ, kemudian bisa diketahui kategori yang tepat untuk data. Intinya, pengguna menentukan kategori terhadap data yang ada.

- CONTEXT-BASED CLASSIFICATION

Jika tadi ada klasifikasi berbasis konten, kali ini ada klasifikasi berbasis konteks. Artinya, proses pengelompokan mengandalkan konteks yang melekat pada data, misalnya lokasi, jenis aplikasi yang digunakan, bahkan pencipta data.

Contoh - contoh algoritma dalam klasifikasi, sebagai berikut

- LOGISTIC REGRESSION

Bisa dibilang, inilah jenis algoritma yang paling dasar dalam proses classification. Meski begitu, keberadaannya sulit untuk digantikan. Dalam algoritma ini, klasifikasi data dilakukan dengan menemukan hubungan antara satu variabel dependen dengan satu (atau lebih) variabel independen.

- K-NEAREST NEIGHBORS

Berikutnya ada algoritma K-nearest neighbors atau lebih sering disebut KNN. Sederhananya, algoritma ini mengikuti prinsip kemiripan data. Data akan dikelompokkan dalam satu kategori yang sama jika ditemukan beberapa kemiripan. K sendiri di sini berarti jumlah “tetangga” atau data di sekitar yang memiliki kemiripan sehingga bisa dipertimbangkan. Untuk bisa memilih K yang tepat, coba jalankan algoritma ini dengan nilai K yang berbeda. Pilihlah nilai dengan kesalahan yang paling minimal.

- **NAIVE BAYES**

Naive Bayes merupakan suatu algoritma yang dibuat berdasarkan Teorema Bayes. Dalam teorema tersebut, diketahui bahwa asumsi harus berubah secara subjektif jika muncul fakta baru. Saat dijadikan dasar algoritma, Naive Bayes berasumsi bahwa kehadiran variabel tertentu pada kelas tidak terkait dengan keberadaan variabel lain. Algoritma ini memiliki tiga classifier, yaitu Bernoulli Naive Bayes, Gaussian Naive Bayes, dan Multinomial Naive Bayes.

- **DECISION TREE**

Decision tree merupakan model yang terinspirasi oleh bagaimana neuron dalam otak manusia bekerja. Tiap neuron pada otak manusia saling berhubungan dan informasi mengalir dari setiap neuron tersebut. Jika diilustrasikan, decision tree memiliki beberapa cabang. Nah, cabang-cabang tersebut mewakili keputusan (decision) yang diambil. Sedangkan, simpul-simpul yang mengaitkan cabang merupakan atribut data. Di sisi lain, hasil keputusan diwakili oleh daun. Hasil tersebut bisa berupa nilai kategoris (untuk klasifikasi data) atau kontinu (regresi data).

- **RANDOM FOREST**

Terakhir ada random forest. Bisa dibilang, algoritma ini merupakan kelanjutan dari algoritma decision tree. Sebab, dalam algoritma ini terdapat kumpulan decision tree untuk membantu proses classification terhadap data. Oleh karenanya, algoritma ini pun dapat digunakan dalam analisis regresi selain dalam classification.

- **NEURAL NETWORK**

Selanjutnya ada algoritma neural network. Disebut demikian karena algoritma ini memang mirip dengan sistem saraf otak manusia. Dalam neural network, terdapat satu set unit input atau output yang saling terkoneksi. Masing-masing koneksi tersebut punya bobot tertentu. Pada fase pembelajaran, koneksi tersebut akan mengupdate bobotnya sehingga mampu memprediksi kategori data.

3. Clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum. Hasil clustering yang baik akan menghasilkan tingkat kesamaan yang tinggi dalam satu kelas dan tingkat kesamaan yang rendah antar kelas. Kesamaan yang dimaksud merupakan pengukuran secara numeric terhadap dua buah objek. Nilai kesamaan antar kedua objek akan semakin tinggi jika kedua objek yang dibandingkan memiliki kemiripan yang tinggi. Begitu juga dengan sebaliknya. Kualitas hasil clustering sangat bergantung pada metode yang dipakai

Ada beberapa jenis clustering yang tersedia yaitu :

- Centroid-based clustering

Clustering berbasis Centroid adalah metode pengelompokan data ke dalam cluster non-hierarki. Cluster jenis ini cenderung lebih efisien, tetapi sensitif terhadap outlier. Tipe ini juga merupakan salah satu algoritma iteratif untuk clustering, dimana cluster terbentuk dari jarak minimum antar titik data ke pusat cluster.

- Density-based Clustering

Pengelompokan berbasis kepadatan menggabungkan wilayah dengan kepadatan yang sama menjadi sebuah grup. Jenis ini menyebabkan masalah dengan data berdimensi tinggi dan berkepadatan tinggi. Metode ini membuat cluster berdasarkan kepadatan setiap titik data. Berikut ini adalah contoh algoritma clustering berbasis centroid DBSCAN (Pengelompokan spasial berbasis kepadatan dari aplikasi bising) OPTICS (order point untuk mengidentifikasi struktur clustering) HDBSCAN (Pengelompokan Spasial Berbasis Kepadatan Hierarki untuk Aplikasi Bising)

- Distribution-based Clustering

Jenis pengelompokan ini mengasumsikan bahwa data terdiri dari distribusi yang mirip dengan distribusi Gaussian. Semakin jauh jarak dari pusat distribusi, semakin kecil kemungkinan titik akan berada di grup distribusi.

- Hierarchical Clustering atau Connectivity based Cluster

Jenis implementasi ini mirip dengan clustering berbasis centroid, yang pada dasarnya mendefinisikan sebuah cluster berdasarkan jarak terpendek antara titik data, kemudian metode ini berfungsi untuk memahami bahwa titik data yang lebih dekat berperilaku sama dengan titik data yang lebih jauh. Pengelompokan titik data direpresentasikan menggunakan dendrogram, Ada beberapa jenis hubungan dendrogram.

4. *Supervised Learning*

Algoritma supervised learning merupakan algoritma machine learning yang proses pembelajarannya di bawah pengawasan guru atau supervisor. Algoritma ini memerlukan

data berlabel untuk membangun sebuah model yang tingkat akurasinya bisa ditingkatkan dari waktu ke waktu. Semakin banyak model tersebut mengolah data, maka tingkat keakurasiannya juga akan semakin tinggi.

Salah satu contoh pengaplikasian supervised learning yang paling menarik adalah memprediksi kondisi cuaca di lokasi tertentu. Untuk membuat prediksi cuaca yang benar, kita perlu memperhitungkan berbagai parameter, termasuk data suhu dari waktu ke waktu, curah hujan, angin, kelembaban, dan lain sebagainya. Metode yang tepat untuk memprediksi suhu adalah metode regresi dengan label output berupa data kontinu, sedangkan metode untuk memprediksi turunnya salju adalah metode klasifikasi binar.

Unsupervised Learning

Algoritma unsupervised learning merupakan proses pembelajaran lebih bebas karena tidak ada pengawasan. Algoritma unsupervised learning lebih bebas dalam proses eksplorasi data karena tidak memiliki data label dan bisa mencari karakteristik data yang tersembunyi. Algoritma ini menggunakan titik data sebagai referensi untuk menemukan struktur dan pola yang ada di dalam data set.

Contoh pengaplikasian algoritma unsupervised learning adalah segmentasi pelanggan, mengurangi kompleksitas suatu masalah, dan memilih fitur yang tepat. Segmentasi pelanggan membutuhkan metode clustering. Metode ini termasuk algoritma unsupervised learning yang bertujuan untuk menemukan kelompok atau cluster alami di dalam data input. Salah satu pendekatan umum dalam clustering adalah membagi titik data sedemikian rupa sehingga setiap titik data yang memiliki kemiripan masuk ke dalam grup yang sama.

5. Usia

$$P(\text{Tua}|\text{Positif}) = (\text{Usia: Tua Berlabel Positif} / \text{Jumlah label Positif})(3/10)=0,3$$

$$P(\text{Tua}|\text{Negatif}) = (\text{Usia: Tua Berlabel Negatif} / \text{Jumlah label Negatif})(2/10)=0,2$$

$$P(\text{Muda}|\text{Positif}) = (\text{Usia: Muda Berlabel Positif} / \text{Jumlah label Positif})(2/10)=0,2$$

$$P(\text{Muda}|\text{Negatif}) = (\text{Usia: Muda Berlabel Negatif} / \text{Jumlah label Negatif})(3/10)=0,3$$

Batuk

$$P(\text{Kering}|\text{Positif}) = (\text{Batuk: Kering Berlabel Positif} / \text{Jumlah label Positif})(4/10)=0,4$$

$$P(\text{Kering} | \text{Negatif}) = (\text{Batuk: Kering Berlabel Negatif} / \text{Jumlah label Negatif})(2/10)=0,2$$

$$P(\text{Basah}|\text{Positif}) = (\text{Batuk: Basah Berlabel Positif} / \text{Jumlah label Positif})(1/10)=0,1$$

$$P(\text{Basah} | \text{Negatif}) = (\text{Batuk: Basah Berlabel Negatif} / \text{Jumlah label Negatif})(3/10)=0,3$$

Suhu Badan

$$P(\text{Tinggi}|\text{Positif}) = (\text{SuhuBadan: Tinggi Berlabel Positif} / \text{Jumlah label Positif})(2/10)=0,2$$

$P(\text{Tinggi} | \text{Negatif}) = (\text{SuhuBadan: Tinggi Berlabel Negatif} / \text{Jumlah label Negatif}) (3/10) = 0,3$

$P(\text{Rendah} | \text{Positif}) = (\text{SuhuBadan: Rendah Berlabel Positif} / \text{Jumlah label Positif}) (3/10) = 0,3$

$P(\text{Rendah} | \text{Negatif}) = (\text{SuhuBadan: Rendah Berlabel Negatif} / \text{Jumlah label Negatif}) (2/10) = 0,2$

Status Corona

$P(\text{Positif}) = (5/10) = 0,5$

$P(\text{Negatif}) = (5/10) = 0,5$

Perhitungan Label Positif (0,006)

Usia(Tua) = 0,3

Batuk(Basah) = 0,1

Suhu Badan(Tinggi) = 0,2

Nilai Positif = $0,3 * 0,1 * 0,2 = 0,006$

Perhitungan Label Negatif (0,018)

Usia(Tua) = 0,2

Batuk(Basah) = 0,3

Suhu Badan(Tinggi) = 0,3

Nilai Negatif = $0,2 * 0,3 * 0,3 = 0,018$

Berdasarkan perhitungan manual algoritma Bayesian Classifier diatas Status Corona yang Paling Tinggi adalah Negatif dengan Hasil 0,018.

6. Diketahui dataset { 4, 6, 10, 12, 3, 20, 40, 33 (nim = 23), 25 } $C1 = \sqrt{((4 - 4)^2 + (6 - 6)^2 + (10 - 10)^2 + (12 - 12)^2 + (3 - 3)^2 + (20 - 20)^2 + (40 - 40)^2 + (33 - 33)^2 + (25 - 25)^2}$
 $= 0$

7. Diketahui data sebagai berikut

Sampel No	X	Y
1	195	72
2	180	53 (nim = 23)
3	178	60

4	189	68
5	192	72
6	198	77

Ambil data ke 1 dan ke 2 sebagai hitungan pertama

Cluster	X	Y
K1	195	72
K2	180	53

Mencari jarak minimum dengan rumus sebagai berikut

$$[(x, y), (a, b)] = \sqrt{(x - a)^2 + (y - b)^2}$$

$$Cluster\ 1 = \sqrt{(195 - 195)^2 + (72 - 72)^2} = 0 \quad Cluster\ 2 = \sqrt{(180 - 180)^2 + (53 - 53)^2} = 0$$

Jarak Cluster 2 ke Cluster 1 (180,53) \rightarrow (195,72)

$$\sqrt{(180 - 195)^2 + (53 - 72)^2} = \sqrt{225 + 361} = \sqrt{586} = 24,20$$

Jarak Cluster 1 ke Cluster 2 (195,72) \rightarrow (180,53)

$$\sqrt{(180 - 195)^2 + (72 - 53)^2} = \sqrt{225 + 361} = \sqrt{586} = 24,20$$

Sehingga didapat hasil sebagai berikut: 26,62

Cluster	Centroid		Kelompok Cluster
	X	Y	
K1	0	24,20	1
K2	26,62	0	2

Menghitung data 3 terhadap cluster 1 (195,72) \rightarrow (178,60)

$$\sqrt{(178 - 195)^2 + (60 - 72)^2} = \sqrt{289 + 144} = \sqrt{433} = 20,8$$

Menghitung data 3 terhadap cluster 2 (180,53) \rightarrow (178,60)

$$\sqrt{(178 - 180)^2 + (60 - 53)^2} = \sqrt{4 + 49} = \sqrt{53} = 7,28$$

Sehingga data 3 masuk ke cluster 2 karena jaraknya lebih sedikit yaitu 7,28

Lalu update nilai centroid pada K2 karena data masuk pada K2

Cluster	X	Y
K1	195	72
K2	$\frac{180 + 178}{2} = 179$	$\frac{53 + 60}{2} = 56$

Menghitung data 4 terhadap cluster 1 (195,72) $\square \rightarrow (189,68)$

$$\sqrt{(189 - 195)^2 + (68 - 72)^2} = \sqrt{36 + 16} = \sqrt{52} = 7,21$$

Menghitung data 4 terhadap cluster 2 yang sudah diperbarui (179,56) $\square \rightarrow (189,68)$

$$\sqrt{(189 - 179)^2 + (68 - 56)^2} = \sqrt{100 + 144} = \sqrt{244} = 15,62$$

Sehingga data 4 masuk ke cluster 1 karena jaraknya lebih sedikit yaitu 7,21

Lalu update nilai centroid pada K1 karena data masuk pada K1

Cluster	X	Y
K1	$\frac{195 + 189}{2} = 192$	$\frac{72 + 68}{2} = 70$
K2	179	56

Menghitung data 5 terhadap cluster 1 yang sudah di update (192,70) $\square \rightarrow (192,72)$

$$\sqrt{(192 - 192)^2 + (72 - 70)^2} = \sqrt{0 + 4} = \sqrt{4} = 2$$

Menghitung data 5 terhadap cluster 2 yang sudah diperbarui (179,55) $\square \rightarrow (192,72)$

$$\sqrt{(192 - 179)^2 + (72 - 56)^2} = \sqrt{13 + 16} = \sqrt{29} = 5,38$$

Sehingga data 5 masuk ke cluster 1 karena jaraknya lebih sedikit yaitu 2

Lalu update nilai centroid pada K1 karena data masuk pada K1

Cluster	X	Y
K1	$\frac{192 + 192}{2} = 192$	$\frac{70 + 72}{2} = 71$
K2	179	56

Menghitung data 6 terhadap cluster 1 yang sudah di update (192,71) $\square \rightarrow (198,77)$

$$\sqrt{(198 - 192)^2 + (77 - 71)^2} = \sqrt{36 + 25} = \sqrt{61} = 7,81$$

Menghitung data 6 terhadap cluster 2 yang sudah diperbarui (179,56) $\square \rightarrow (198,77)$

$$\sqrt{(198 - 179)^2 + (77 - 56)^2} = \sqrt{361 + 441} = \sqrt{802} = 28,31$$

Sehingga data 6 masuk ke cluster 1 karena jaraknya lebih sedikit yaitu 7,81

Dari hasil perhitungan hingga dataset ke 6 didapat hasil sebagai berikut

Sampel No	X	Y	Cluster/Klmpk
1	195	72	1
2	180	53	2
3	178	60	2
4	189	68	1
5	192	72	1
6	198	77	1

8. Permasalahan : Memprediksi Kelayakan Kredit

Latar Belakang

Kredit merupakan memperoleh barang atau pinjaman uang yang pembayarannya dilakukan di kemudian hari dengan cicilan atau angsuran sesuai dengan perjanjian. Masalah kredit tidak pernah lepas dari peran bank sebagai lembaga keuangan. Bahkan, pemberian kredit merupakan kegiatan utama bank sebagai lembaga keuangan. Pemberian kredit kepada nasabah adalah kegiatan rutin yang mempunyai resiko tinggi. Dalam pelaksanaannya, kredit yang bermasalah (kredit macet) sering terjadi akibat analisis kredit yang tidak hati-hati atau kurang cermat dalam proses pemberian kredit, maupun karakter nasabah yang tidak baik. Untuk mencegah terjadinya kredit macet, seorang analis kredit perbankan harus mampu mengambil keputusan yang tepat untuk menerima ataupun menolak pengajuan kredit. Pemberian kredit memiliki resiko kredit macet atau non performing loan (NPL)

yang dapat berpengaruh pada kegiatan utama bank. Menurut media publikasi Statistik Perbankan Indonesia yang diterbitkan oleh Departemen Perizinan dan Informasi Perbankan, rasio kredit macet pada perbankan di Indonesia pada tahun 2017 adalah 4,06% dan menurun di tahun 2018 menjadi 1,9%. Pada tahun 2019 kredit macet meningkat drastis menjadi 13,22%. Sedangkan pertumbuhan kredit perbankan Indonesia pada tahun 2017 sebesar 8,24% dan pada tahun 2018 pertumbuhan kredit meningkat 11,75%. Untuk tahun 2019, pertumbuhan kredit menurun menjadi 6,08% (www.ojk.go.id).

Algoritma Naives Bayes

Algoritma Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang ditemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Algoritma Naïve Bayes menggunakan metode probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. untuk membantu analisis kreditmeningkatkan kehati-hatian dalam pengambilan keputusan pemberian kredit, dibutuhkan suatu aplikasi untuk memprediksi penilaian layak atau tidaknya pemohon kredit ketika analisa kredit berdasarkan data dan informasi kredit-kredit sebelumnya. Maka algoritma yang cocok adalah Algoritma Naïve Bayes.

Dataset dari permasalahan

Nama	Jenis Kelamin	Umur	Pekerjaan	Jumlah Pinjaman	Jangka Waktu	Jaminan	Penghasilan	Status
Ahmad Tofa	L	40-49	Wiraswasta	100.000.000-149.999.999	24	Sertifikat Rumah	0 – 10jt	Lancar
Ananda Syarif	L	30-39	Karyawan Swasta	100.000.000-149.999.999	36	BPKB MB	5jt – 10jt	Macet
Ali Rahmawan	L	21-29	Karyawan Swasta	150.000.000-200.000.000	36	BPKB MB	0 – 5jt	Lancar
Aisyah Rahmawati	P	40-49	Wiraswasta	5.000.000-49.999.999	12	Sertifikat Rumah	0 – 5jt	Lancar
Allya Cantika	P	30-39	PNS	5.000.000-49.999.999	12	Sertifikat Rumah	0 – 5jt	Macet
Fendy Ahmad	L	40-49	Karyawan Swasta	100.000.000-149.999.999	24	Sertifikat Rumah	5jt – 10jt	Lancar
Dimas Satrio	L	30-39	Karyawan Swasta	100.000.000-149.999.999	24	BPKB MB	0 – 5jt	Lancar
Dina Putri	P	21-29	PNS	50.000.000-99.999.999	12	Sertifikat Rumah	0 – 5jt	Lancar

Fatkur Setiawan	L	40-49	Karyawan Swasta	100.000.000-149.999.999	24	BPKB MB	0 – 5jt	Macet
Galuh Putra	P	30-39	Wiraswasta	100.000.000-149.999.999	24	BPKB MB	0 – 5jt	Macet

Hasil data testing

Nama	Jenis Kelamin	Umur	Pekerjaan	Jumlah Pinjaman	Jangka Waktu	Jaminan	Penghasilan	Status
Hasibuan Ilham	L	30-39	Karyawan Swasta	100.000.000-149.999.999	24	BPKB MB	5jt – 10jt	?

Dengan rumus Naïve Bayes dapat dihitung probabilitasnya sebagai berikut:

a. Menghitung jumlah variabel kelas/label

$$-P(Y=Lancar) = 6/10 = 0,6$$

$$-P(Y=Macet) = 4/10 = 0,4$$

b. Menghitung jumlah kasus yang sama dengan kelas yang sama

$$-P(\text{Jenis Kelamin} = L \mid Y=Lancar) = 4 / 6 = 0,67$$

$$-P(\text{Jenis Kelamin} = L \mid Y=Macet) = 2 / 4 = 0,5$$

$$-P(\text{Umur} = 30-39 \mid Y=Lancar) = 1 / 6 = 0,167$$

$$-P(\text{Umur} = 30-39 \mid Y=Macet) = 3 / 4 = 0,75$$

$$-P(\text{Pekerjaan} = swasta \mid Y=Lancar) = 3 / 6 = 0,5$$

$$-P(\text{Pekerjaan} = swasta \mid Y=Macet) = 2 / 4 = 0,5$$

$$-P(\text{Jumlah Pinjaman} = 100.000.000 – 149.999.999 \mid Y=Lancar) = 3 / 6 = 0,5$$

$$-P(\text{Jumlah Pinjaman} = 10.000.000 – 149.999.999 \mid Y=Macet) = 3 / 4 = 0,75$$

$$-P(\text{jangka waktu pengembalian} = 12 bulan \mid Y=Lancar) = 3 / 6 = 0,5$$

$$-P(\text{jangka waktu pengembalian} = 12 bulan \mid Y=Macet) = 2 / 4 = 0,5$$

$$-P(\text{Jaminan} = BPKB MB \mid Y=Lancar) = 2 / 6 = 0,33$$

$$-P(\text{Jaminan} = BPKB MB \mid Y=Macet) = 3 / 4 = 0,75$$

$$-P(\text{Penghasilan} = 5jt-10jt \mid Y=Lancar) = 1 / 6 = 0,167$$

$$-P(\text{Penghasilan} = 5jt-10jt \mid Y=Macet) = 1 / 4 = 0,25$$

c. Kalikan semua hasil labellancar dan Macet

$$-P(\text{Jenis Kelamin} = L \mid Y=Lancar) \times P(\text{Umur} = 30-39 \mid Y=Lancar) \times P(\text{Pekerjaan} = swasta \mid Y=Lancar) \times P(\text{Jumlah Pinjaman} = 100.000.000 – 149.999.999 \mid Y=Lancar) \times P(\text{jangka waktu pengembalian} = 24 bulan \mid Y=Lancar) \times P(\text{Jaminan} = BPKB MB \mid Y=Lancar) \times P(\text{Penghasilan} = 5jt-10jt \mid Y=Lancar) \times P(Y=Lancar) = 0,67 \times 0,167 \times 0,5 \times 0,5 \times 0,5 \times 0,33 \times 0,167 \times 0,6 = 0,0004624693425$$

- $P(\text{Jenis Kelamin} = L \mid Y = \text{Macet}) \times P(\text{Umur} = 30-39 \mid Y = \text{Macet}) \times P(\text{Pekerjaan} = \text{swasta} \mid Y = \text{Macet}) \times P(\text{Jumlah Pinjaman} = 100.000.000 - 149.999.999 \mid Y = \text{Macet}) \times P(\text{jangka waktu pengembalian} = 24 \text{ bulan} \mid Y = \text{Macet}) \times P(\text{Jaminan} = \text{BPKB MB} \mid Y = \text{Macet}) \times P(\text{Penghasilan} = 5\text{jt}-10\text{jt} \mid Y = \text{Macet}) \times P(Y = \text{Macet}) = 0,5 \times 0,75 \times 0,5 \times 0,75 \times 0,5 \times 0,75 \times 0,25 \times 0,4 = 0,0052734375$

d. Bandingkan hasil kelas lancar dan Macet

Dari hasil diatas, terlihat bahwa nilai probabilitas tertinggi ada pada kelas/label $(P \mid Y = \text{Macet})$ sehingga dapat disimpulkan bahwa status calon nasabah tersebut termasuk golongan “Macet”

Nama	Jenis Kelamin	Umur	Pekerjaan	Jumlah Pinjaman	Jangka Waktu	Jaminan	Penghasilan	Status
Hasibuan Ilham	L	30-39	Karyawan Swasta	100.000.000-149.999.999	24	BPKB MB	5jt – 10jt	Macet

Referensi

<https://ejournal.unib.ac.id/index.php/pseudocode/article/download/3852/2092>

<https://journal.unnes.ac.id/nju/index.php/jte/article/view/8585/5674>

<http://repository.unmuhjember.ac.id/470/1/Jurnal.pdf>

9. N

10. A

11. A

12. A