



# Penerapan Data Mining Untuk Memprediksi Besarnya Pembayaran Pajak Kendaraan Pada: (Badan Pengelolaan Pajak Dan Retribusi Daerah Upt Samsat Medan Selatan) Menggunakan Algoritma Expectation Maximization

Suhada<sup>1</sup>, Guidio L Ginting<sup>2</sup>, Rivalri K Hondro<sup>3</sup>

Program Studi Teknik Informatika, STMIK Budi Darma, Medan, Indonesia

Email: [suhada@gmail.com](mailto:suhada@gmail.com)

## INFORMASI ARTIKEL

*Sejarah Artikel:*  
 Diterima Redaksi : 07 Juni 2021  
 Revisi Akhir : 15 Juli 2021  
 Diterima : 20 Juli 2021  
 Diterbitkan Online : 28 Juli 2021

## KATA KUNCI

Kata Kunci : Pajak, Algoritma Expectation

## KORESPONDENSI

E-mail: [suhada@gmail.com](mailto:suhada@gmail.com)

## A B S T R A C T

*Pajak adalah Melakukan salah satu peran dan tugas fiskus dalam diterapkannya sistem pemungutan self-assessment di Indonesia. Masalahnya adalah Pemahaman dan kesadaran masyarakat untuk patuh bayar pajak masih minim. Hal ini karena pengetahuan mengenai pajak yang di peroleh masyarakat belum optimal dan masyarakat belum memahi dalam memprediksi berapa besar pajak kendaraannya. Oleh karena itu masyarakat perlu diberikan pemahaman pajak supaya kesadaran dapat terbangun.*

*Algoritma Expectation Maximization merupakan salah satu algoritma yang di gunakan untuk klasifikasi atau pengelompokan data. Dalam algoritma ini, ada dua hal yang digunakan secara bergantian yaitu E step yang menghitung nilai ekspektasi dari likelihood termasuk latent variabel seolah-olah seperti mereka ada, dan M step menghitung nilai estimasi ML dari parameter dengan memaksimalkan nilai ekspektasi dari likelihood yang ditemukan pada E step.*

*Sumatera utara merupakan provinsi dengan jumlah penduduk terpadat antar provinsi di sumatera selatan urutan di indonesia. Namun potensi yang cukup besar akan penerimaan pajak masih belum banyak tersentuh. Padahal semakin besar sumber pendapatan yang berassal dari potensi daerah dan bukan dari bantuan pemerintah pusat, maka daerah semakin mampu untuk lebih mensukseskan pembangunan didaerah sesuai kebutuhan masyarakatnya.*

## 1. PENDAHULUAN

Pajak adalah Melakukan salah satu peran dan tugas fiskus dalam diterapkannya sistem pemungutan self-assessment di Indonesia. Definisi pemeriksaan menurut pasal 1 ayat (25) UU No. 28 Tahun 2007 adalah serangkaian kegiatan menghimpun dan mengolah data, keterangan, atau bukti yang dilaksanakan secara objektif dan profesional berdasarkan suatu standar pemeriksaan untuk menguji kepatuhan pemenuhan kewajiban perpajakan dan/atau untuk tujuan lain dalam rangka melaksanakan ketentuan peraturan perundang-undangan perpajakan. Pasal 3 ayat (1) Peraturan Menteri Keuangan No. 199/PMK.03/2007 menyebutkan, “Ruang lingkup pemeriksaan untuk menguji kepatuhan pemenuhan kewajiban perpajakan dapat meliputi satu, beberapa, atau seluruh jenis pajak, baik untuk satu atau beberapa masa pajak, bagian tahun pajak, atau tahun pajak dalam tahun-tahun lalu maupun tahun berjalan.

Masalahnya adalah Pemahaman dan kesadaran masyarakat untuk patuh bayar pajak masih minim. Hal ini karena pengetahuan mengenai pajak yang di peroleh masyarakat belum optimal dan masyarakat belum memahi dalam memprediksi berapa besar pajak kendaraannya. Oleh karena itu masyarakat perlu diberikan pemahaman pajak supaya kesadaran dapat terbangun.

Algoritma Expectation Maximization merupakan salah satu algoritma yang di gunakan untuk klasifikasi atau pengelompokan data. Dalam algoritma ini, ada dua hal yang digunakan secara bergantian yaitu E step yang menghitung nilai ekspektasi dari likelihood termasuk latent variabel seolah-olah seperti mereka ada, dan M step menghitung nilai estimasi ML dari parameter dengan memaksimalkan nilai ekspektasi dari likelihood yang ditemukan pada E step. Sumatera utara merupakan provinsi dengan jumlah penduduk terpadat antar provinsi di sumatera selatan urutan di indonesia. Namun potensi yang cukup besar akan penerimaan pajak masih belum banyak tersentuh. Padahal semakin besar sumber pendapatan yang berassal dari potensi daerah dan bukan dari bantuan pemerintah pusat, maka daerah semakin mampu untuk lebih mensukseskan pembangunan didaerah sesuai kebutuhan masyarakatnya

## 2. METODOLOGI PENELITIAN

### 2.1 Data Mining

Data mining adalah cara pandang dan pengetahuan yang berbeda membuat para ahli memberikan definisi berbeda tentang data mining. Sebagian para ahli menyatakan bahwa data mining adalah langkah analisis terhadap proses penemuan pengetahuan didalam basisdata atau knowledge discovery in databases yang di singkat KDD(Fayyad et al. 1996). Pengetahuan bisa berupa pola data atau relasi antar data yang valid (yang tidak di ketahui sebelumnya).[1] DataMining merupakan gabungan sejumlah disiplin ilmu komputer, yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari artificial intelligence, machine learning, statistics, dan database systems.

### 2.2 Pajak

Pengertian pajak dan pandangan para ahli dalam bidang tersebut memberikan berbagai definisi tentang pajak yang berbeda-beda, tetapi pada dasarnya definisi tersebut mempunyai tujuan yang sama. Berdasarkan definisi pajak menurut Undang-Undang Republik Indonesia Nomor 28 tahun 2007 tentang perubahan ketiga atas Undang- Undang Nomor 6 tahun 1983 mengenai Ketentuan Umum dan Tata Cara Perpajakan, pajak adalah kontribusi wajib kepada Negara yang terutang oleh orang pribadi atau badan yang bersifat memaksa berdasarkan Undang-Undang, dengan tidak mendapatkan imbalan secara langsung dan digunakan untuk keperluan negara bagi sebesar-besarnya kemakmuran rakyat. Standar pemeriksaan (fungsional) pajak per bulan adalah 8 SPP (Surat Perintah Pemeriksaan) pajak. Berdasarkan hasil wawancara yang diperoleh dari KPP, pemeriksaan sering mengalami penurunan dari standar yang telah ditentukan. Hal ini menunjukkan bahwa masih terdapat celah dalam pemeriksaan pajak. Secara tidak langsung hal ini menunjukkan bahwa pemeriksa pajak harus mampu meningkatkan kinerjanya.

### 2.3 Algoritma Expectation Maximization

Algoritma Expectation Maximization sering digunakan untuk menemukan nilai estimasi Maximum Likelihood (ML) dari parameter dalam sebuah model probabilistic, dimana model juga tergantung pada latent variabel yang belum diketahui. Dalam algoritma ini, ada dua hal yang digunakan secara bergantian yaitu E step yang menghitung nilai ekspektasi dari likelihood termasuk latent variabel seolah-olah seperti mereka ada, dan M step menghitung nilai estimasi ML dari parameter dengan memaksimalkan nilai ekspektasi dari likelihood yang ditemukan pada E step. Algoritma EM adalah sebuah metode optimisasi iteratif untuk estimasi Maksimum Likelihood (ML) yang berguna dalam permasalahan data yang tidak lengkap (incomplete data). Kasus khusus dimana algoritma EM digunakan untuk memprediksi rata-rata populasi dan varians tidak diketahui dan harus diperkirakan mempunyai tahap Ekspektasi (Expectation Step) dan tahap Maksimisasi (Maximization Step).

Tahap Ekspektasi atau Expectation Step (E Step) 1. Hitung nilai parameter dari data yang ada.

$$\tilde{\mu} = \frac{1}{n} \sum_{j=1}^n x_{jk} \quad k = 1, 2, \dots, p$$

$$\tilde{\sigma}_k = \tilde{\sigma}_{kk} = \frac{1}{n} \sum_{j=1}^n (x_{jk} - \bar{x}_k)^2 \quad k = 1, 2, \dots, p$$

$$\tilde{\sigma}_{ik} = \frac{1}{n} \sum_{j=1}^n (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k) \quad i = 1, 2, \dots, p \quad k = 1, 2, \dots, p$$

Dengan

$$\tilde{\mu} = \text{rata-rata mean}$$

$$\tilde{\sigma}_k = \text{varians}$$

$$\tilde{\sigma}_{ik} = \text{kovarians}$$

2. Masukan ke persamaan

Untuk setiap persamaan  $X_j^{(1)}$  adalah komponen yang hilang, dan  $X_j^{(2)}$  adalah komponen yang ada. Untuk memprediksi  $\tilde{\mu}$  dan  $\tilde{\Sigma}$  digunakan mean distribusi

$$\begin{aligned} \tilde{x}_j^{(1)} &= E\left(\tilde{X}_j^{(1)} \mid x_j^{(2)}; \tilde{\mu}, \tilde{\Sigma}\right) \\ &= \tilde{\mu}^{(1)} + \tilde{\Sigma}_{12} \tilde{\Sigma}_{22}^{-1} (x_j^{(2)} - \tilde{\mu}^{(2)}) \end{aligned}$$

Memprediksi kontribusi  $x_j^{(1)}$  untuk  $T_1$  :

$$\begin{aligned} \overline{x_j^{(1)} x_j^{(1)}} &= E\left(X_j^{(1)} X_j^{(1)} \mid x_j^{(2)}; \tilde{\mu}, \tilde{\Sigma}\right) \\ &= \tilde{\Sigma}_{11} - \tilde{\Sigma}_{12} \tilde{\Sigma}_{22}^{-1} \tilde{\Sigma}_{21} + \tilde{X}_j^{(1)} \tilde{X}_j^{(1)} \\ \overline{x_j^{(1)} x_j^{(2)}} &= E\left(X_j^{(1)} X_j^{(2)} \mid x_j^{(2)}; \tilde{\mu}, \tilde{\Sigma}\right) \\ &= \tilde{x}_j^{(1)} \tilde{x}_j^{(2)} \end{aligned}$$

Memprediksi kontribusi  $x_j^{(1)}$  untuk  $T_2$  :

bersyarat  $X^{(1)}$  diberikan  $X^{(2)}$

$$\begin{aligned} \tilde{T}_1 &= \begin{bmatrix} \tilde{x}_{11} + x_{21} + x_{31} + \tilde{x}_{41} \\ x_{12} + x_{22} + x_{32} + \tilde{x}_{42} \\ x_{13} + x_{23} + x_{33} + x_{43} \end{bmatrix} \\ \tilde{T}_2 &= \begin{bmatrix} \overline{x_{11}^2 + x_{21}^2 + x_{31}^2 + x_{41}^2} & & \\ \frac{x_{11}x_{12} + x_{21}x_{22} + x_{31}x_{32} + x_{41}x_{42}}{x_{11}x_{13} + x_{21}x_{23} + x_{31}x_{33} + x_{41}x_{43}} & \frac{x_{12}^2 + x_{22}^2 + x_{32}^2 + x_{42}^2}{x_{12}x_{13} + x_{22}x_{23} + x_{32}x_{33} + x_{42}x_{43}} & \\ & & x_{13}^2 + x_{23}^2 + x_{33}^2 + x_{43}^2 \end{bmatrix} \end{aligned}$$

Kontribusi pertama dijumlahkan untuk setiap  $j$   $x$  dengan komponen yang hilang. Hasil ini digabungkan dengan data sampel menghasilkan 1  $T$  dan 2  $T$  Menentukan matriks 1  $T$  dan

$$\begin{aligned} \tilde{T}_1 &= \begin{bmatrix} \tilde{x}_{11} + x_{21} + x_{31} + \tilde{x}_{41} \\ x_{12} + x_{22} + x_{32} + \tilde{x}_{42} \\ x_{13} + x_{23} + x_{33} + x_{43} \end{bmatrix} \\ \tilde{T}_2 &= \begin{bmatrix} \overline{x_{11}^2 + x_{21}^2 + x_{31}^2 + x_{41}^2} & & \\ \frac{x_{11}x_{12} + x_{21}x_{22} + x_{31}x_{32} + x_{41}x_{42}}{x_{11}x_{13} + x_{21}x_{23} + x_{31}x_{33} + x_{41}x_{43}} & \frac{x_{12}^2 + x_{22}^2 + x_{32}^2 + x_{42}^2}{x_{12}x_{13} + x_{22}x_{23} + x_{32}x_{33} + x_{42}x_{43}} & \\ & & x_{13}^2 + x_{23}^2 + x_{33}^2 + x_{43}^2 \end{bmatrix} \end{aligned}$$

Tahap Maksimisasi atau *Maximization Step* (M Step)

$$\begin{aligned} \tilde{\mu} &= \frac{\tilde{T}_1}{n} \\ \tilde{\Sigma} &= \frac{1}{n} \tilde{T}_2 - \tilde{\mu} \tilde{\mu}' \end{aligned}$$

Uji 2  $\chi^2$  adalah pengujian hipotesis mengenai perbandingan antara frekuensi observasi / yang benar-benar terjadi dengan frekuensi harapan / ekspektasi. Nilai 2  $\chi^2$  adalah nilai kuadrat. Oleh karena itu nilai 2  $\chi^2$  selalu positif. Uji 2  $\chi^2$  digunakan untuk menunjukkan apakah ada pengaruh data hilang terhadap nilai awal dalam sebuah data dan kemudian dibandingkan dengan nilai hitung dengan rumus :

$$n(\tilde{\mu} - \mu)' \tilde{\Sigma}^{-1} (\tilde{\mu} - \mu) \leq \chi_p^2(\alpha)$$

### 3. ANALISA DAN PEMBAHASAN

#### 3.1 Analisa

Masalah yang terjadi adalah Pemahaman dan kesadaran masyarakat untuk patuh bayar pajak masih minim. Hal ini karena pengetahuan mengenai pajak yang di peroleh masyarakat belum optimal dan masyarakat belum memahi dalam memprediksi berapa besar pajak kendaraannya. Oleh karena itu masyarakat perlu diberikan pemahaman pajak supaya kesadaran dapat terbangun. Pada proses ini akan dilakukan penerapan data mining untuk mempermudah masyarakat dalam memprediksi besarnya pajak kendaraan, Sekaligus menyadarkan masyarakat akan wajib pajak secara tepat waktu. Penelitian ini digunakan dengan bantuan pemrograman Visual Basic.Net dan menggunakan database microsoft access.

Algoritma Expectation Maximization merupakan salah satu algoritma yang digunakan untuk klasifikasi atau pengelompokan data. Dalam algoritma ini, ada dua hal yang digunakan secara bergantian yaitu E step yang menghitung nilai ekspektasi dari likelihood termasuk latent variabel seolah-olah seperti mereka ada, dan M step menghitung nilai estimasi ML dari parameter dengan memaksimalkan nilai ekspektasi dari likelihood yang ditemukan pada E step. Contoh kasus yang menggunakan algoritma expectation maximization pada data jumlah pajak kendaraan di UPT Samsat Medan Selatandapat dilihat pada tabel di bawah ini :

**Tabel 3.1** Realisasi Penerimaan Pendapatan Asli Daerah

Realisasi Penerimaan Pendapatan Asli Daerah (PAD)			
	Bulan&Tahun	Fitur(X)	Jenis Penerimaan
Y=1	Desember Tahun 2016	X1	Pajak Kendaraan Bermotor(PKB)
		X2	BBN Kendaraan Bermotor
		X3	Pajak Air Permukaan(APU)
		X4	Pendapatan Denda Pajak
		X5	Denda Pajak Kendaraan Bermotor
		X6	Sumbangan Pihak Ketiga

Target (Y)

**Tabel 3.2** Realisasi Penerimaan Pendapatan Asli Daerah Lanjutan

	Bulan&Tahun	Fitur(X)	Jenis Penerimaan
Y2	Desember Tahun 2017	X7	Pajak Kendaraan Bermotor(PKB)
		X8	BBN Kendaraan Bermotor
		X9	Pajak Air Permukaan(APU)
		X10	Pendapatan Denda Pajak
		X11	Denda Pajak Kendaraan Bermotor
		X12	Sumbangan Pihak Ketiga

Terdapat kesamaan jenis penerimaan dengan bulan dan tahun yang berbeda ang sudah tercantum dalam PAD.

Perhitungan Ekspektasi probabilitas dilakukan dengan contoh 2 kelas Y=1, Dan Y=2 dengan 4 fitur X=1, X=2, X=3, X=4. Kelas Y adalah target yang kemungkinana dicapai. Fitur X merupakan representasi dari jenis penerimaan dan nilainya. Sebagai contoh jika pajak kendaraan bermotor bernilai 20, 10, 5, 0. Fitur bernilai 1 pada nilai yang dimiliki PAD dan bernilai 0 jika tidak memiliki nilai. Berikut adalah contoh perhitungan dalam proses pengolahan pajak dan retribusi daerah UPT Samsat Medan Selatan dengan menggunakan algoritma *Expectation-Maximization*. Perhitungan dengan algoritma ini mempunyai dua tahap utama yakni tahap perkiraan (*Expectation*) dan tahap maksimalisasi (*Maximization*), di mana tahap ini akan diulang (*increment*) sampai nilai kelas Y mencapai nilai 0 atau nilai 1. a. *Expectation Step*

Pada tahap ini, akan menghitung probabilitas objek terhadap kluster.

**Tabel 3.3** Kondisi Awal Probabilitas Target Y yang di miliki

Y=1	Y=2	X1	X2	X3	X4
?	?	0	1	0	1
?	?	1	0	1	0
?	?	0	1	1	1
?	?	1	1	1	0
?	?	0	0	1	1

Pada tabel 3.3 di atas, kelas Y=1, Y=2 dan Y=3 adalah data awal yang akan dihitung probabilitas target yang dimiliki oleh setiap fitur X1, X2, X3 dan X4. Tabel tersebut menunjukkan bahwa nilai target Y belum diketahui, untuk data setiap fitur dapat dilihat bernilai 1 jika jenis penerimaan yang diambil misalnya bernilai 20 dan 0 jika jenis penerimaan tersebut bukan bernilai 20.

**Tabel 3.4** Frekuensi Fitur X

Y	X1	X2	X3	X4
1	?	?	?	?
2	?	?	?	?

Tabel 3.4 di atas adalah table frekuensi yang dimiliki setiap fitur terhadap kelas Y=1, dan Y=2. Pada awal kondisi nilai frekuensi setiap fitur x1, x2, x3 dan x4 belum diketahui. Untuk dapat menghitung nilai frekuensi setiap fitur terhadap setiap target, maka digunakan sebuah model parameter. Tabel 3.5 berikut ini menampilkan nilai model parameter.

**Tabel 3.5** Model Parameter

Y	X1	X2	X3	X4
1	0.1	0.7	0.8	0.8
2	0.9	0.3	0.2	0.2

Pada tabel 3.5 yang berisi model parameter masing-masing nilai X1, X2, X3 dan X4 diperoleh dari secara random dimana jumlah total seluruh fitur yang sama dari seluruh fitur harus berjumlah 1 (satu). Dapat dilihat bahwa  $x1 \rightarrow 0.1+0.9= 1$ ,  $X2 \rightarrow 0.3+0.7 =1$ ,  $X3 \rightarrow 0.8 + 0.2 =1$ ,  $X4 \rightarrow 0.8+0.2=1$  dan  $\alpha \rightarrow 0.1+0.2+0.1=1$ . Dari nilai yang dihasilkan secara random ini berikutnya akan dihitung frekuensi setiap fitur X beserta nilai probabilitas terhadap target Y seperti yang ditampilkan pada tabel 3.6 berikut ini.

**Tabel 3.6** Frekuensi fitur X baru dan Probabilitas target Y baru

Prob Y=1					
X1	X2	X3	X4	P(y)	N
0.9	0.3	0.8	0.8	0.7	0.03024
0.1	0.7	0.2	0.2	0.7	0.00784
0.9	0.3	0.8	0.8	0.7	0.12096
0.1	0.7	0.8	0.2	0.7	0.00336
0.9	0.7	0.2	0.2	0.7	0.00756
Prob Y=2					
X1	X2	X3	X4	P(y)	N
0.8	0.3	0.9	0.1	0.2	0.00432
0.2	0.7	0.1	0.9	0.2	0.00252
0.8	0.3	0.1	0.1	0.2	0.00048

0.2	0.3	0.1	0.9	0.2	0.00108
0.8	0.7	0.1	0.9	0.2	0.03888

Perhitungan setiap fitur X baru pada tabel4.5 diperoleh dari nilai model parameter yang terdapat di tabel3.6. Selanjutnya untuk mengisi frekuensi X baru pada tabel4.5 tersebut akan di periksa nilai awal fitur X yang terdapat pada tabel4.2. Perhitungan yang digunakan adalah jika nilai fitur X pada data awa l bernilai 1, maka ambil nilai frekuensi yang dihasilkan secara random pada tabel model parameter, jika 0 maka frekuensi X baru adalah 1 dikurangi nilai frekuensi yang dihasilkan secara random pada model parameter tabel4.4. Selanjutnya setelah semua nilai frekuensi fitur X baru dihasilkan, maka dapat dihitung nilai probabilitas target Y baru.

**Tabel 3.7** Nilai Probabilitas baru dari target Y

Y=1	Y=2
baru	baru
0.85	0.12
0.55	0.18
1.00	0.00
0.48	0.16
0.13	0.69

Maka probabilitas setiap instance/ fitu terhadap sebuah target Y adalah sebagai berikut:  $\prod P(y|k) = P'(y) / \sum P'(y)$ .

Di mana nilai  $\prod P(y|k)$  adalah nilai probabilitas setiap fitur terhadap sebuah target Y, nilai  $P'(y)$  adalah nilai probabilitas yang dimiliki sebuah target Y, nilai  $\sum P'(y)$  adalah jumlah seluruh probabilitas yang dimiliki seluruh target Y.

- a.  $\prod P(y|k)$  pada target Y=1
  - Instance1 :  $0.03024 / (0.03024 + 0.00432) = 0.85$
  - Instance2 :  $0.00784 / (0.00784 + 0.00252) = 0.55$
  - Instance3 :  $0.12096 / (0.12096 + 0.00048) = 1.00$
  - Instance4 :  $0.00336 / (0.00336 + 0.00432) = 0.48$
  - Instance5 :  $0.00756 / (0.00756 + 0.03888) = 0.13$
- b.  $\prod P(y|k)$  pada target Y=2
  - Instance1 :  $0.00432 / (0.03024 + 0.00432) = 0.12$
  - Instance2 :  $0.00252 / (0.00784 + 0.00252) = 0.18$
  - Instance3 :  $0.00048 / (0.12096 + 0.00048) = 0.00$
  - Instance4 :  $0.00432 / (0.00336 + 0.00432) = 0.16$
  - Instance5 :  $0.03888 / (0.00756 + 0.03888) = 0.69$

Tabel 4.6 di atas menunjukkan nilai probabilitas target Y yang baru. Nilai tersebut ditunjukkan oleh nilai Y=1 baru dan nilai Y=2 baru. Nilai tersebut diperoleh dari hasil perbandingan setiap nilai N dengan total nilai N. Dengan perhitungan sebagaiberikut:

$$Y \text{ baru} = N(\text{prob } Y1) / (N(\text{prob } Y1) + N(\text{prob } Y2))$$

Selanjutnya nilai baru tersebut akan digunakan pada iterasi pertama sebagai tahap Maximization.

Berikut tahap Maximization nilai probabilitas setiap target Y:

**Tabel 3.8** Data Probabilitas Target yang dimiliki setiap fitur dalam iterasi -I

Target		Fitur			
Y=1	Y=2	X1	X2	X3	X4
0.85	0.12	0	1	0	1
0.55	0.18	1	0	1	0
1.00	0.00	0	1	1	1
0.48	0.16	1	1	1	0
0.13	0.69	0	0	1	1

**Tabel 3.9** Tabel Frekuensi nilai dalam iterasi -I

Y	X1	X2	X3	X4	Total
1	1.04	2.46	3.03	2.84	4.01
2	0.33	0.97	0.34	0.13	1.15

Probabilitas seluruh data (X1.....X4) terhadap kelas  $Y=\{0,1,2\}$ : a.  $Y'(X)$

Untuk  $Y=1$ :

$$Y'(x) = \sum P(X|Y) = 0.85 + 0.55 + 1.00 + 0.48 + 0.13 + 1.00 = 4.01$$

b.  $Y'(X)$  Untuk  $Y=2$ :

$$Y'(x) = \sum P(X|Y) = \{0.12 + 0.18 + 0.00 + 0.16 + 0.69 + 0.00\} = 1.15$$

Jadi, masing –masing nilai fitur x adalah hasil jumlah seluruh nilai fitur x dikali dengan nilai probabilitas target Y yang dihasilkan oleh hasil tahap sebelumnya. Selanjutnya, pada iterasi ini lakukan kembali perhitungan model parameter.

**Tabel 3.10** Model Parameter dalam iterasi -I

Y	X1	X2	X3	X4	Alpha
1	0.26	0.61	0.75	0.71	0.67
2	0.29	0.84	0.30	0.11	0.19

Untuk nilai model parameter yang dihasilkan dalam iterasi I tahap Maximization ini dilakukan dengan cara membagi nilai frekuensi setiap fitur x dengan total frekuensi fitur. Hingga selanjutnya seluruh tahap yang dilakukan dalam proses maximization ini dilakukan secara berulang terus menerus hingga nilai probabilitas target yang dihasilkan mencapai nilai 0 atau 1 yang artinya setiap bulan dapat ditentukan akan termasuk kepada jenis penerimaan apa yang cocok (target Y yang terpilih) berdasarkan nilai jenis penerimaan yang diperoleh (fitur x). Berikut adalah contoh nilai probabilitas target Y yang sudah mencapai nilai ambang 0 atau 1 .

**Tabel 3.11.** Nilai probabilitas Iterasi Terakhir dari Target Y (Targetnya bernilai 0 atau 1)

Y=0	Y=1	Y=2	X1	X2	X3	X4
0.00	1.00	0.00	0	1	0	1
0.00	0.00	1.00	1	0	1	0
1.00	0.00	0.00	0	1	1	1
0.00	1.00	1.00	1	1	1	0
1.00	1.00	0.00	0	0	1	1

#### 4. KESIMPULAN

Dari analisa dan pembahasan yang dilakukan diatas, maka dapat diambil kesimpulan:

1. Setelah memprediksi besarnya pajak kendaraan maka pengguna kendaraan telah mudah untuk mengetahui besarnya biaya pajak kendaraan meraka.
2. Proses untuk memprediksi pajak kendaraan menggunakan pemrograman Matlab telah dapat digunakan.
3. Setelah merancang pemograman untuk memprediksi pajak maka penerapan Algoritma Expectation Maximization telah dapat diterapkan dalam pemograman matlab .

#### REFERENCES

- [1] M. S. Dr. Suyanto, S.T., Data Mining Untuk Klasifikasi Dan Klasterisasi Data. Bandung: Informatika Bandung, 2017.
- [2] L. E. T. Kusriani, Algoritma Data Mining. Yogyakarta: C.V ANDI OFFSET, 2009.
- [3] Mardiasmo, “Efek Tivitas Dan Kontribusi Penerimaan Pajak Hotel Dan Restoran Terhadap Pad Kora Manado,” 2013.
- [4] M. K. Novianti Dian, Analisa Dan Perancangan Sistem Informasi Berorientasi Objek. Yogyakarta: CV BUDI UTAMA, 2014.
- [5] Gunaidi Abdia Away. (2014). The Shortcut of matlab programming. Bandung Informatika Bandung