# PERBANDINGAN METODE KLASIFIKASI *NAIVE* BAYES DAN *K-NEAREST NEIGHBOR* PADA DATA STATUS PEMBAYARAN PAJAK PERTAMBAHAN NILAI DI KANTOR PELAYANAN PAJAK PRATAMA SAMARINDA ULU

**SKRIPSI**

**Fatihah Noor Ramaulidyah**
**NIM. 1607015025**

**PROGRAM STUDI STATISTIKA**
**JURUSAN MATEMATIKA**
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**
**UNIVERSITAS MULAWARMAN**
**SAMARINDA**
**2020**

# PERBANDINGAN METODE KLASIFIKASI *NAIVE* BAYES DAN *K-NEAREST NEIGHBOR* PADA DATA STATUS PEMBAYARAN PAJAK PERTAMBAHAN NILAI DI KANTOR PELAYANAN PAJAK PRATAMA SAMARINDA ULU

**SKRIPSI**

**Diajukan Kepada**

**Program Studi Statistika Jurusan Matematika**
**Fakultas Matematika dan Ilmu Pengetahuan Alam**
**Universitas Mulawarman untuk memenuhi sebagian persyaratan**
**memperoleh gelar Sarjana Sains bidang Ilmu Statistika**

**Oleh :**

**Fatihah Noor Ramaulidyah**
**NIM. 1607015025**

**PROGRAM STUDI STATISTIKA**
**JURUSAN MATEMATIKA**
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**
**UNIVERSITAS MULAWARMAN**
**SAMARINDA**
**2020**

## HALAMAN PENGESAHAN

**Skripsi Sarjana** berjudul **Perbandingan Metode Klasifikasi *Naive* Bayes dan *K-Nearest Neighbor* pada Data Status Pembayaran Pajak Pertambahan Nilai di Kantor Pelayanan Pajak Pratama Samarinda Ulu** oleh **Fatihah Noor Rahmaulidyah** telah dipertahankan di depan Dewan Penguji pada tanggal 14 Mei 2020.

## SUSUNAN TIM PEMBIMBING

Menyetujui,

| Pembimbing I, | Pembimbing II, |
|---|---|
| **Memi Noor Hayati, M.Si.** | **Rito Goejantoro, M.Si.** |
| **NIP. 19880503 201404 2 001** | **NIP. 19720626 200003 1 001** |

Mengetahui,
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Mulawarman

**Dr. Eng. Idris Mandang, M.Si.**
**NIP. 19711008 199802 1 001**

## PERNYATAAN KEASLIAN SKRIPSI

Dengan ini saya menyatakan bahwa dalam Skripsi yang berjudul "**Perbandingan Metode Klasifikasi *Naive* Bayes dan *K-Nearest Neighbor* pada Data Status Pembayaran Pajak Pertambahan Nilai di Kantor Pelayanan Pajak Pratama Samarinda Ulu**" tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu perguruan tinggi manapun. Sepanjang pengetahuan saya, tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Demikian pernyataan ini dibuat dengan sebenar-benarnya. Saya sanggup menerima konsekuensi akademik dikemudian hari apabila pernyataan yang dibuat ini tidak benar.

Samarinda, 14 Mei 2020

Fatihah Noor Rahmaulidyah

# ABSTRAK

Klasifikasi adalah pengelompokan sistematis dari suatu obyek ke dalam kelompok atau golongan tertentu berdasarkan ciri yang sama. Metode klasifikasi yang digunakan pada penelitian ini adalah *naive* Bayes dan *K-Nearest Neighbor* yang memiliki tingkat akurasi relatif tinggi. Penelitian ini bertujuan untuk melakukan perbandingan tingkat akurasi klasifikasi pada data status pembayaran pajak pertambahan nilai (PPN). Data yang digunakan adalah data wajib pajak badan di KPP Pratama Samarinda Ulu Tahun 2018 dengan status pembayaran PPN patuh atau tidak patuh dan menggunakan 3 variabel bebas yaitu pendapatan, bentuk badan dan status pelaporan pajak. Pengukuran akurasi menggunakan APER pada metode *naive* Bayes sebesar 17,07% sedangkan pada metode K-NN sebesar 19,51%. Hasil perbandingan pengukuran akurasi dari kedua metode tersebut menunjukkan bahwa metode *naive* Bayes memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan metode K-NN.



**Kata kunci**: klasifikasi, *naive* Bayes, *k-nearest neighbor*, pajak.

## *ABSTRACT*

*Classification is a systematic grouping of objects into certain groups based on the same characteristics. The classification method used in this research were naive Bayes and K-Nearest Neighbor which had a relatively high degree of accuracy. This research aimed to compare the level of classification accuracy on the status data of value-added tax (VAT) payment. The data used is data on corporate taxpayers at Samarinda Ulu Tax Office in 2018 with the status of VAT payment being compliant or non-compliant and used 3 independent variables are income, type of business entity and tax reporting status. Measurement of accuracy used APER in the Naive Bayes method was 17.07% and in K-Nearest Neighbor method was 19,51%. The comparison results of accuracy measurements between the two methods showed that the naive Bayes method had a higher level of accuracy than the K-Nearest Neighbor method.*



***Keywords****: classification, naive Bayes, k-nearest neighbor, tax.*

# KATA PENGANTAR

Segala puji dan syukur penulis panjatkan atas kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul "**Perbandingan Metode Klasifikasi *Naive* Bayes dan *K-Nearest Neighbor* pada Data Status Pembayaran Pajak Pertambahan Nilai di Kantor Pelayanan Pajak Pratama Samarinda Ulu**".

Tugas akhir ini ditulis untuk memenuhi syarat dalam menyelesaikan pendidikan Strata Satu (S1) pada Program Studi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman. Penulis menyadari selama penyusunan tugas akhir ini tidak terlepas dari bantuan, bimbingan dan motivasi dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Bapak Dr. Eng. Idris Mandang, M.Si selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman.
2. Bapak Dr. Suyitno, S.Pd., M.Sc selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman.
3. Ibu Dr. Sifriani, S.Pd., M.Si selaku Ketua Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman.
4. Ibu Memi Nor Hayati, M.Si selaku Dosen Pembimbing I dan Bapak Rito Goejantoro, M.Si selaku Dosen Pembimbing II yang telah banyak memberikan bimbingan, arahan, saran dan motivasi kepada penulis sehingga tugas akhir ini dapat terselesaikan.
5. Bapak Dr. Syaripuddin, M.Si selaku Dosen Penguji I dan Ibu Meiliyani Siringoringo, M.Si selaku Dosen Penguji II yang telah memberikan arahan, kritik dan masukan yang membangun demi kesempurnaan tugas akhir ini.
6. Kedua orang tua Ayah Muhammad Noor dan Ibu Ellinah telah memberikan doa, motivasi, serta dukungan moril dan materil kepada penulis.
7. Kakak-kakak tercinta (Fachrul Noor Rizal, Faizal Noor Ramadhan, Amelia Besari, Reza Gadis, Eny Agustiandini, Suci Arti) yang selalu memberi motivasi dan semangat kepada penulis.

8. Sahabat-sahabat saya (Mirdiansyah, Putri Yunita Sari, Feni Indah Rianti, Nana Nirwana, Imasdiani) yang selalu menemani serta memberikan dukungan dan motivasi kepada penulis selama proses penyusunan tugas akhir.
9. Teman seperjuangan (Amanah, Widya Suerni, Reni Pratiwi, Riska Veronika) yang saling membantu dan memberi masukan satu sama lain.
10. Teman-teman Statistika angkatan 2016 yang telah memberikan dukungan dari teman satu ke teman yang lainnya selama penyusunan tugas akhir ini.

Penulis menyadari bahwa dalam menulis tugas akhir ini masih banyak terdapat kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari pembaca. Akhirnya dengan segala hormat dan kerendahan hati penulis berharap semoga penulisan tugas akhir ini dapat menambah wawasan dan pengetahuan serta bermanfaat bagi pembaca.

Samarinda, 14 Mei 2020

Fatihah Noor Rahmaulidyah

## DAFTAR ISI

Halaman

**HALAMAN JUDUL** ..... i
**HALAMAN PENGESAHAN** ..... ii
**HALAMAN PERNYATAAN KEASLIAN SKRIPSI** ..... iii
**ABSTRAK** ..... iv
***ABSTRACT*** ..... v
**KATA PENGANTAR** ..... vi
**DAFTAR ISI** ..... viii
**DAFTAR TABEL** ..... ix
**DAFTAR GAMBAR** ..... x
**DAFTAR SIMBOL** ..... xi
**DAFTAR LAMPIRAN** ..... xii
**BAB 1 PENDAHULUAN**
1.1 Latar Belakang ..... 1
1.2 Batasan Masalah ..... 3
1.3 Rumusan Masalah ..... 3
1.4 Tujuan Penelitan ..... 4
1.5 Manfaat Penelitian ..... 4
**BAB 2 TINJAUAN PUSTAKA**
2.1 Data *Mining* ..... 5
2.1.1 Teknik Data *Mining* ..... 5
2.1.2 Kelompok Data *Mining* ..... 6
2.2 Klasifikasi ..... 7
2.2.1 Metode *Naive* Bayes ..... 8
2.2.2 Metode *K-Nearest Neighbor* ..... 12
2.3 Data *Training* dan Data *Testing* ..... 14
2.4 Pengukuran Tingkat Akurasi ..... 15
2.5 Pajak Pertambahan Nilai ..... 16
2.5.1 Dasar Hukum Pajak Pertambahan Nilai ..... 16

2.5.2 Subjek Pajak Pertambahan Nilai ..... 17
2.5.3 Objek Pajak Pertambahan Nilai ..... 17
2.5.4 Tarif Pajak Pertambahan Nilai ..... 17
**BAB 3 METODOLOGI PENELITIAN**
3.1 Waktu dan Tempat Penelitian ..... 19
3.2 Rancangan Penelitian ..... 19
3.3 Teknik Sampling ..... 19
3.4 Populasi dan Sampel Penelitian ..... 19
3.5 Teknik Pengambilan Data ..... 20
3.6 Variabel Penelitian ..... 20
3.7 Teknik Analisis Data ..... 21
3.8 Kerangka Penelitian ..... 25
**BAB 4 HASIL DAN PEMBAHASAN**
4.1 Analisis Statistika Deskriptif ..... 26
4.2 Membagi Data *Training* dan Data *Testing* ..... 29
4.3 Klasifikasi *Naive* Bayes ..... 30
4.4 Klasifikasi *K-Nearest Neighbor* ..... 37
4.5 Perbandingan Tingkat Akurasi Klasifikasi ..... 42
**BAB 5 PENUTUP**
5.1 Kesimpulan ..... 44
5.2 Saran ..... 44
**DAFTAR PUSTAKA** ..... 46
**LAMPIRAN** ..... 48
**RIWAYAT HIDUP** ..... 89

## DAFTAR TABEL

Halaman

**Tabel 2.1** Tabel Klasifikasi ..... 15
**Tabel 3.1** Variabel Penelitian ..... 20
**Tabel 4.1** Probabilitas *Prior* Pada Setiap Kelompok ..... 31
**Tabel 4.2** Probabilitas Variabel Pendapatan Pada Setiap Kelompok ..... 32
**Tabel 4.3** Probabilitas Variabel Bentuk Badan Pada Setiap Kelompok ..... 33
**Tabel 4.4** Probabilitas Variabel Status Pelaporan Pada Setiap Kelompok ..... 34
**Tabel 4.5** Hasil Klasifikasi Metode *Naive* Bayes ..... 36
**Tabel 4.6** Nilai APER Pada Metode *Naive* Bayes untuk Setiap Percobaan Randomisasi Data ..... 37
**Tabel 4.7** Ranking Jarak Euclid Antara Data *Testing* Pertama dan Data *Traning* ..... 37
**Tabel 4.8** Hasil Klasifikasi Metode K-NN ..... 40
**Tabel 4.9** Nilai APER Pada Metode K-NN untuk Seluruh Nilai *K* ..... 41
**Tabel 4.10** Nilai APER Pada Metode K-NN untuk Setiap Percobaan Randomisasi Data ..... 42
**Tabel 4.11** Hasil Pengukuran Tingkat Akurasi Pada Metode *Naive* Bayes dan K-NN ..... 42

## DAFTAR GAMBAR

Halaman

**Gambar 3.1** Diagram alir metode *naive* Bayes ..... 22
**Gambar 3.2** Diagram alir metode K-NN..... 24
**Gambar 3.3** Kerangka penelitian..... 25
**Gambar 4.1** Persentase status pembayaran pajak..... 26
**Gambar 4.2** Persentase pendapatan wajib pajak badan..... 27
**Gambar 4.3** Persentase bentuk badan..... 28
**Gambar 4.4** Persentase status pelaporan pajak ..... 29

## DAFTAR SIMBOL

| Simbol | Arti |
|---|---|
| $N$ | Jumlah seluruh data (data *training* + data *testing*) |
| $P(A \mid B)$ | Peluang $A$ dengan syarat $B$ terjadi |
| $P(B \cap A)$ | Peluang $A$ dan $B$ terjadi |
| $P(A)$ | Peluang terjadinya $A$ |
| $P(B)$ | Peluang terjadinya $B$ |
| $P(Y)$ | Peluang munculnya kelompok $Y$ |
| $X$ | Variabel bebas |
| $Y_1$ | Kelompok pertama |
| $Y_2$ | Kelompok kedua |
| $q$ | Banyaknya kelompok |
| $n$ | Banyaknya data *training* |
| $n^*$ | Banyaknya data *testing* |
| $p$ | Banyaknya variabel bebas |
| $K$ | Nilai parameter K-NN |
| $d$ | Jarak Euclid |
| $x_{ik}$ | Nilai variabel bebas ke-$k$ dari data *training* ke-$i$ |
| $x^*_{jk}$ | Nilai variabel bebas ke-$k$ dari data *testing* ke-$j$ |
| $y_i$ | Nilai variabel terikat dari data *training* ke-$i$ |
| $y_j$ | Nilai variabel terikat dari data *testing* ke-$j$ |

## DAFTAR LAMPIRAN

**Lampiran 1.** Data Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 ..... 29
**Lampiran 2.** *Syntax Pie Chart* dengan Program R ..... 54
**Lampiran 3.** *Syntax* dan *Output* Randomisasi Data dengan Program R ..... 55
**Lampiran 4.** Data Randomisasi Percobaan Keempat Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 ..... 58
**Lampiran 5.** Data *Traning* Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 ..... 64
**Lampiran 6.** Data *Testing* Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 ..... 69
**Lampiran 7.** *Syntax Naive* Bayes dengan Program R ..... 71
**Lampiran 8.** Nilai *Posterior* untuk Setiap Kelompok dan Hasil Klasifikasi dengan Metode *Naive* Bayes ..... 78
**Lampiran 9.** *Syntax K-Nearest Neighbor* dengan Program R ..... 80
**Lampiran 10.** Perhitungan Jarak Euclid antara Data *Traning* dan Data *Testing* ..... 85
**Lampiran 11.** Hasil Klasifikasi dengan Metode *K-Nearest Neighbor* ..... 87
**Lampiran 12.** Surat Pengantar Permohonan Izin Pengambilan Data ..... 89
**Lampiran 13.** Surat Permohonan Izin Pengambilan Data ..... 90

# BAB 1
# PENDAHULUAN

## 1.1 Latar Belakang

*Data mining* merupakan proses yang meliputi pengumpulan dan pemakaian informasi penting untuk menemukan pola dalam data yang berukuran besar (Han dan Kamber, 2006). Perkembangan *data mining* tidak terlepas dari kemajuan teknologi informasi yang memungkinkan data dalam jumlah yang besar dapat terakumulasi. Menurut Larose (2005), *data mining* dibagi menjadi beberapa kelompok berdasarkan tugas atau pekerjaan yang dapat dilakukan yaitu deskripsi, estimasi, prediksi, klastering, asosiasi, dan klasifikasi.

Klasifikasi adalah pengelompokan sistematis dari suatu obyek ke dalam kelompok atau golongan tertentu berdasarkan ciri yang sama (Towa dan Tairas, 2002). Dalam klasifikasi terdapat dua pekerjaan utama yang dilakukan, yaitu pembangunan model untuk disimpan sebagai memori dan penggunaannya untuk melakukan pengenalan atau prediksi pada suatu obyek data lain. Terdapat banyak algoritma klasifikasi antara lain *naive* Bayes, *Decision Tree*, *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM) dan *K-Nearest Neighbor*.

Pada penelitian ini penulis akan membandingkan dua algoritma klasifikasi yaitu *naive* Bayes dan *K-Nearest Neighbor*. Adapun dipilihnya metode *naive* Bayes dan K-NN sebagai metode klasifikasi yang dibandingkan adalah karena tingkat akurasi dari kedua metode ini relatif tinggi. Hal ini dibuktikan pada penelitian sebelumnya oleh Antaristi (2017) tentang aplikasi klasifikasi penentuan pengajuan kartu kredit menggunakan metode *naive* Bayes (Studi Kasus: Bank BNI Syariah Surabaya). Hasil penelitian ini mendapatkan persentase keakuratan sebesar 98,67%. Sementara itu penelitian oleh Prahudaya (2017) mengenai metode klasifikasi mutu jambu biji menggunakan K-NN berdasarkan fitur warna dan tekstur mampu memberikan akurasi terbaik pada $K = 3$ yaitu sebesar 91,25%. Penelitian terkait lainnya oleh Eko (2014) yaitu tentang perbandingan metode klasifikasi *naive* Bayes dan K-NN pada analisis data status kerja di Kabupaten Demak mendapatkan nilai ketepatan klasifikasi dengan metode *naive*

Bayes sebesar 94,09% sedangkan ketepatan klasifikasi menggunakan metode K-NN sebesar 96,06%. Penelitian selanjutnya oleh Indriati (2015) tentang perbandingan kinerja metode *naive* Bayes dan K-NN dalam pengkategorian buku komputer menunjukkan bahwa rata-rata akurasi pada metode K-NN mencapai 96% dan *naive* Bayes mencapai 98%.

Klasifikasi *naive* Bayes merupakan suatu metode klasifikasi yang berakar pada teorema Bayes yaitu dengan memprediksi peluang di masa depan berdasarkan masa lalu (Bustami, 2013). Menurut Kursini (2009), keuntungan dari pendekatan *naive* Bayes yaitu hanya memerlukan *training* data yang kecil untuk mengestimasi parameter yang dibutuhkan untuk klasifikasi.

Sedangkan klasifikasi *K-Nearest Neighbor* (K-NN) termasuk dalam kategori *lazy learner* yang berdasar pada perbandingan *K* tetangga terdekat. Menurut Prasetyo (2012), Algoritma K-NN merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi atau jarak suatu data dengan data yang lain. Algoritma K-NN merupakan salah satu algoritma paling populer serta memiliki keuntungan yaitu dapat diproses dengan mudah dan sederhana (Harrington, 2012).

Metode klasifikasi *naive* Bayes maupun K-NN dapat diaplikasikan dalam berbagai bidang, salah satunya adalah bidang perpajakan. Pajak merupakan penghasilan negara yang tujuan pembayarannya digunakan untuk membiayai pembangunan yang berguna bagi kepentingan bersama. Pembayaran pajak bukan hanya merupakan kewajiban kenegaraan tetapi merupakan hak bagi setiap warga negara untuk berpartisipasi dalam pembiayaan negara (Judisseno, 2005). Bagi negara pajak merupakan sumber pendapatan atau instrumen yang primer dan strategis. Dengan adanya pajak, pemerintah dapat melaksanakan pembangunan, melangsungkan kinerja pemerintahan, serta menyejahterakan masyarakat.

Pada umumnya pajak terbagi menjadi dua yaitu pajak langsung dan pajak tidak langsung. Pajak langsung contohnya adalah Pajak Penghasilan (PPh), sedangkan pajak tidak langsung contohnya adalah Pajak Pertambahan Nilai (PPN). Menurut Mardiasmo (2009), sebagai salah satu jenis pajak yang menjadi sumber pemasukan negara, PPN lebih menonjol dalam meningkatkan penerimaan

negara jika dibandingkan dengan PPh. Hal tersebut dikarenakan hampir seluruh barang-barang kebutuhan hidup masyarakat merupakan hasil produksi yang terkena PPN, dengan kata lain sebagian transaksi di bidang perdagangan, industri dan jasa pada prinsipnya terkena PPN.

Berbagai kebijakan pemerintah di bidang penerimaan negara yang telah dilakukan diarahkan pada upaya peningkatan penerimaan pajak. Pemerintah melalui *self assesment* memberikan kepercayaan kepada wajib pajak untuk menghitung sendiri besar PPN terhutangnya, menyetorkan ke bank persepsi dan kemudian melaporkan secara teratur ke Kantor Pelayanan Pajak dalam bentuk surat pemberitahuan. Kantor Pelayanan Pajak bertugas dan berfungsi sebagai pengumpulan, pencarian dan pengolahan data, pelaksana pendaftaran wajib pajak, pendataan objek dan subjek pajak, serta peyajian informasi pajak. Dalam pendataan objek dan subjek pajak diperlukan adanya suatu sistem klasifikasi untuk mempermudah penyajian informasi pajak yang dibutuhkan. Berdasarkan latar belakang tersebut, maka penulis tertarik untuk melakukan penelitian ilmiah dengan judul "Perbandingan Metode Klasifikasi *Naive* Bayes dan *K-Nearest Neighbor* pada Data Status Pembayaran Pajak Pertambahan Nilai di Kantor Pelayanan Pajak Pratama Samarinda Ulu".

## 1.2 Batasan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka batasan masalah pada penelitian ini adalah sebagai berikut :

1. Nilai parameter $K$ yang digunakan dalam K-NN adalah $K=1$, $K=3$, $K=5$, $K=7$, dan $K=9$.
2. Proses klasifikasi menggunakan proporsi 80% data *training* dan 20% data *testing*.
3. Pengukuran tingkat akurasi dari hasil klasifikasi metode *naive* Bayes dan K-NN menggunakan nilai *Apparent Error Rate* (APER).

## 1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana hasil klasifikasi status pembayaran pajak pertambahan nilai di wilayah Samarinda Ulu dengan menggunakan metode *naive* Bayes?
2. Bagaimana hasil klasifikasi status pembayaran pajak pertambahan nilai di wilayah Samarinda Ulu dengan menggunakan metode K-NN?
3. Bagaimana perbandingan hasil klasifikasi status pembayaran pajak pertambahan nilai dengan menggunakan metode *naive* Bayes dan K-NN berdasarkan tingkat akurasi?

## 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut :

1. Untuk memperoleh hasil klasifikasi status pembayaran pajak pertambahan nilai di wilayah Samarinda Ulu dengan menggunakan metode *naive* Bayes.
2. Untuk memperoleh hasil klasifikasi status pembayaran pajak pertambahan nilai di wilayah Samarinda Ulu dengan menggunakan metode K-NN.
3. Untuk memperoleh perbandingan hasil klasifikasi status pembayaran pajak pertambahan nilai dengan menggunakan metode *naive* Bayes dan K-NN berdasarkan tingkat akurasi.

## 1.5 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memberikan informasi bahwa metode *naive* Bayes dan metode K-NN dapat digunakan menjadi salah satu alternatif pengklasifikasian.
2. Memberikan pengetahuan tentang pengaplikasian metode *naive* Bayes dan K-NN terhadap kehidupan sehari-hari terutama pada bidang perpajakan.
3. Menjadi bahan referensi bagi penelitian selanjutnya.

# BAB 2
# TINJAUAN PUSTAKA

## 2.1 Data *Mining*

Data *mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial yang tersimpan dalam *database* besar (Turban, 2005).

Menurut Vulandari (2017), data *mining* adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. Dari beberapa definisi di atas dapat ditarik kesimpulan bahwa data *mining* merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi-informasi yang nantinya dapat digunakan.

### 2.1.1 Teknik Data *Mining*

Menurut Vulandari (2017) teknik yang digunakan dalam data *mining* erat kaitannya dengan "penemuan" (*discovery*) dan "pembelajaran" (*learning*) yang terbagi dalam tiga metode utama pembelajaran yaitu:

a. *Supervised Learning*

   *Supervised Learning* adalah teknik yang paling sering digunakan. Teknik ini sama dengan "*programming by example*". Teknik ini melibatkan fase pelatihan dimana data pelatihan historis yang karakter-karakternya dipetakkan ke hasil-hasil yang telah diketahui diolah dalam algoritma data *mining*. Proses ini melatih algoritma untuk mengenali variabel-variabel dan nilai-nilai kunci yang nantinya akan digunakan sebagai dasar dalam membuat perkiraan-perkiraan ketika diberikan data baru.

b. *Unsupervised Learning*

   Teknik pembelajaran ini tidak melibatkan fase pelatihan seperti yang terdapat

pada *supervised learning*. Teknik ini bergantung pada penggunaan algoritma yang mendeteksi semua pola seperti *associations* yang muncul dari kriteria penting yang spesifik dari data masukan. Pendekatan ini mengarah pada pembuatan banyak aturan (*rules*) yang mengkarakterisasikan penemuan seperti *associations*, *clusters* dan *segments*. Aturan-aturan ini kemudian dianalisis untuk menemukan hal-hal penting.

c. *Reinforcement Learning*

Teknik pembelajaran ini jarang digunakan dibandingkan dengan dua teknik lainnya, namun memiliki penerapan-penerapan yang terus dioptimalkan dari waktu ke waktu dan memiliki kontrol adaptif. *Reinforcement learning* sangat tepat digunakan untuk menyelesaikan masalah-masalah yang sulit serta bergantung pada waktu.

### 2.1.2 Kelompok Data Mining

Menurut Larose (2005), data *mining* dibagi menjadi beberapa kelompok berdasarkan tugas atau pekerjaan yang dapat dilakukan, yaitu:

1. Deskripsi

   Terkadang peneliti dan analisis mencoba menemukan cara untuk mendeskripsikan pola dan *trend* yang tersembunyi dalam data.

2. Estimasi

   Estimasi adalah menerka variabel terikat yang dipengaruhi oleh variabel bebas. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi di mana *record* adalah kumpulan dari elemen-elemen data yang saling terkait dalam sebuah basis data.

3. Prediksi

   Prediksi hampir sama dengan estimasi, nilai hasil estimasi ada di masa sekarang sedangkan nilai hasil prediksi mungkin akan ada di masa yang mendatang.

4. *Cluster*

   *Clustering* merupakan pengelompokan *record*, pengamatan atau memperhatikan untuk membentuk kelompok obyek-obyek yang memiliki

kemiripan. *Cluster* adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record* dalam *cluster* yang lain.

5. Asosiasi

   Tugas asosiasi dalam data *mining* adalah untuk menemukan variabel yang muncul dalam satu waktu. Salah satu implementasi dari asosiasi adalah *market basket analysis* atau analisis keranjang belanja.

6. Klasifikasi

   Pada klasifikasi terdapat variabel target sebagai kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah.

## 2.2 Klasifikasi

Klasifikasi adalah pengelompokan yang sistematis dari obyek, gagasan, buku dan benda-benda lain ke dalam kelompok atau golongan tertentu berdasarkan ciri-ciri yang sama. Klasifikasi dalam kehidupan sehari-hari sudah banyak dilakukan seperti di supermarket, pasar, toko buku dan pedagang yang mengelompokan barang dagangannya yang sejenis dalam satu kelompok yang sama (Towa dan Tairas, 2002).

Menurut Prasetyo (2014), klasifikasi merupakan suatu pekerjaan menilai obyek data untuk memasukkannya ke dalam kelompok tertentu dari sejumlah kelompok yang tersedia. Dalam klasifikasi terdapat dua pekerjaan utama yang dilakukan, yaitu pembangunan model sebagai prototipe untuk disimpan sebagai memori dan penggunaannya untuk melakukan pengenalan/klasifikasi/prediksi pada suatu obyek data lain agar diketahui di kelompok mana obyek data tersebut dalam model yang sudah disimpannya. Sedangkan menurut Johnson dan Wichern (2007), klasifikasi disebut juga sebagai metode pengalokasian dengan tujuan untuk memilih atau memasukkan pengamatan (obyek baru) ke dalam kelompok yang telah mempunyai label kelompok.

Berdasarkan cara pelatihannya, algoritma-algoritma klasifikasi dapat dibagi menjadi dua macam, yaitu *eager learner* dan *lazy learner*. Algoritma-algoritma

yang masuk dalam kategori *eager learner* didesain untuk melakukan pembacaan, pelatihan, dan pembelajaran pada data *training* untuk dapat memetakan dengan benar setiap vektor masukan ke label kelas keluarannya, sehingga di akhir proses pelatihan model sudah dapat melakukan pemetaan dengan benar. Proses prediksi dilakukan menggunakan model yang tersimpan di dalam memori, sehingga dengan cara ini prediksi berjalan dengan cepat, namun harus dibayar dengan proses pelatihan yang lama. Algoritma-algoritma yang masuk dalam kategori ini diantaranya adalah *Decision Tree*, *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM), dan Bayesian (Prasetyo, 2014).

Sementara algoritma-algoritma yang masuk dalam kategori *lazy learner* hanya sedikit melakukan pelatihan (atau bahkan tidak sama sekali). Cara ini hanya menyimpan sebagian data atau seluruh data *training*, kemudian menggunakannya ketika proses prediksi. Hal ini menyebabkan proses prediksi menjadi lama karena model harus membaca kembali semua data *training* untuk dapat memberikan keluaran label kelas dengan benar pada data *testing*. Kelebihan algoritma seperti ini adalah proses pelatihan berjalan dengan cepat. Algoritma-algoritma yang masuk dalam kategori ini diantaranya adalah *rote classifier*, *K-Nearest Neighbor* (K-NN), *Fuzzy K-Nearest Neighbor* (FK-NN) dan regresi linier (Prasetyo, 2014).

### 2.2.1 Metode *Naive* Bayes

Algoritma *naive Bayes* adalah salah satu algoritma yang terdapat pada teknik data *mining* klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuan Inggris yaitu Thomas Bayes, *naive Bayes* memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya, sehingga dikenal dengan Teorema Bayes. Teorema tersebut dikombinasikan dengan *naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya (Bustami, 2013).

Definisi lain mengatakan *naive* Bayes merupakan sebuah pengklasifikasi probabilistik sederhana yang menghitung sekumpulan peluang dengan menjumlahkan frekuensi dan kombinasi nilai dari data *set* yang diberikan. *Naive*

Bayes didasarkan pada asumsi penyederhanaan bahwa nilai variabel secara kondisional saling bebas jika diberikan nilai *output*. Dengan kata lain jika diberikan nilai *output*, peluang mengamati secara bersama adalah produk dari peluang individu (Riduwan, 2013).

Keuntungan penggunaan metode *naive* Bayes adalah metode ini hanya membutuhkan jumlah data *training* yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena diasumsikan sebagai variabel independen, maka hanya varian dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians (Santosa, 2017).

Menurut Prasetyo (2012), persamaan dari Teorema Bayes pada umumnya adalah sebagai berikut:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{2.1}$$

dengan:

$P(A|B) =$ peluang terjadinya $A$ dengan syarat $B$ telah terjadi

$P(B|A) =$ peluang terjadinya $B$ dengan syarat $A$ telah terjadi

$P(A)$ = peluang terjadinya $A$

$P(B)$ = peluang terjadinya $B$

Pada proses klasifikasi diperlukan sejumlah petunjuk yaitu variabel-variabel untuk menentukan kelompok mana yang tepat bagi obyek yang dianalisis tersebut sehingga teorema Bayes disesuaikan sebagai berikut:

$$P(Y|X_1, X_2, ..., X_p) = \frac{P(Y)P(X_1, X_2, ..., X_p|Y)}{P(X_1, X_2, ..., X_p)} \tag{2.2}$$

dengan:

$P(Y|X_1, X_2, ..., X_p) =$ peluang masuknya obyek dengan karakteristik variabel tertentu dalam kelompok $Y$ (*posterior*)

$P(X_1, X_2, ..., X_p|Y) =$ peluang kemunculan variabel-variabel pada obyek yang masuk kelompok $Y$ (*likelihood*)

$P(Y)$ = peluang munculnya kelompok $Y$ sebelum masuknya obyek (*prior*)

$P(X_1, X_2, ..., X_p)$ = peluang kemunculan variabel-variabel pada obyek secara umum (*evidence*)

Pada persamaan (2.2) dapat juga dituliskan secara sederhana sebagai berikut:

$$Posterior = \frac{Prior \times Likelihood}{Evidence} \tag{2.3}$$

Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai *posterior* kelompok lainnya untuk menentukan kelompok suatu obyek akan diklasifikasikan. Mengklasifikasikan suatu obyek dapat ditentukan dengan memilih kelompok yang memiliki *posterior* terbesar, nilai *evidence* selalu tetap untuk setiap kelompok pada satu sampel yaitu bernilai 1 dan merupakan pembagi pada setiap kelompok sehingga dalam perhitungan *posterior* hanya cukup mengalikan nilai *prior* dengan *likelihood*. Nilai prior yang merupakan peluang munculnya kelompok $Y$ sebelum masuknya obyek dapat dihitung menggunakan persamaan sebagai berikut:

$$P(Y_g) = \frac{n_g}{N} \tag{2.4}$$

dengan:

$P(Y_g)$ = peluang munculnya kelompok $Y$ ke-$g$, $g = 1, 2, .., q$

$n_g$ = banyaknya pengamatan pada kelompok ke-$g$

Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $P(Y \mid X_1, X_2, ..., X_p)$ menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned} P(Y \mid X_1, X_2, ..., X_p) &= P(Y)P(X_1 \mid Y)P(X_2, X_3, ..., X_p \mid Y, X_1) \\ &= P(Y)P(X_1 \mid Y)P(X_2, X_3, ..., X_p \mid Y, X_1) \\ &\quad P(X_3, X_4, ..., X_p \mid Y, X_1 X_2) \\ &= P(Y)P(X_1 \mid Y)P(X_2 \mid Y, X_1)P(X_3 \mid Y, X_1 X_2)... \\ &\quad P(X_4, X_5, ..., X_p \mid Y, X_1, X_2, X_3) \end{aligned} \tag{2.5}$$

$$= P(Y)P(X_1 \mid Y)P(X_2 \mid Y, X_1)P(X_3 \mid Y, X_1, X_2)...$$
$$P(X_p \mid Y, X_1, X_2, X_3, ..., X_{p-1})$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai peluang yang hampir mustahil untuk dianalisa satu persatu. Kompleksnya faktor-faktor syarat yang mempengaruhi nilai peluang menyebabkan perhitungan tersebut menjadi sulit untuk dilakukan, maka digunakan asumsi independensi yang sangat tinggi (*naive*), bahwa masing-masing petunjuk $(X_1, X_2, ..., X_p)$ saling bebas (*independent*) satu sama lain sehingga berlaku suatu kesamaan sebagai berikut:

$$P(X_a \mid X_b) = \frac{P(X_a \cap X_b)}{P(X_b)} = \frac{P(X_a)P(X_b)}{P(X_b)} = P(X_a) \tag{2.6}$$

untuk $a \neq b$ sehingga $P(X_a \mid Y, X_b) = P(X_a \mid Y)$

Pada Persamaan (2.6) dapat disimpulkan bahwa asumsi independensi *naive* tersebut membuat syarat peluang menjadi sederhana sehingga perhitungan menjadi mungkin dilakukan. Selanjutnya penjabaran $P(Y \mid X_1, X_2, ..., X_p)$ dapat disederhanakan menjadi:

$$\begin{aligned} P(Y \mid X_1, X_2, ..., X_p) &= P(Y)P(X_1 \mid Y)P(X_2 \mid Y)P(X_3 \mid Y)...P(X_p \mid Y) \\ &= P(Y)\prod_{k=1}^{p} P(X_k \mid Y) \end{aligned} \tag{2.7}$$

Persamaan (2.7) merupakan model dari teorema *naive* Bayes yang selanjutnya akan digunakan dalam proses klasifikasi. Pada umumnya teorema *naive* Bayes mudah dihitung untuk nilai pengamatan variabel bebas bertipe kategorik, namun untuk nilai pengamatan variabel bebas dengan tipe numerik (*non* kategorik) ada perlakuan khusus sebelum diproses menggunakan *naive* Bayes yaitu dengan cara sebagai berikut:

a. Melakukan diskritisasi pada setiap nilai pengamatan variabel bebas kontinu dan mengganti nilai pengamatan tersebut dengan nilai interval diskrit. Pendekatan ini dilakukan dengan mentransformasi ke dalam skala ordinal.

b. Mengasumsikan bentuk tertentu dari distribusi peluang untuk nilai pengamatan kontinu dan memperkirakan parameter distribusi dengan data *training*.

(John dan Langley, 1995).

### 2.2.2 Metode *K-Nearest Neighbor*

*K-Nearest Neighbor* (K-NN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari titik *query* yang baru diklasifikasikan berdasarkan mayoritas dari label kelas pada K-NN. Tujuan dari algoritma K-NN adalah mengklasifikasi objek baru berdasarkan atribut dan *training* data (Larose, 2005). Klasifikasi tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori yang menggunakan jumlah terbanyak diantara klasifikasi dari nilai $K$ sebagai prediksi dari titik *query* yang baru.

Metode klasifikasi K-NN memiliki beberapa tahap, yang pertama nilai parameter $K$ yang merupakan jumlah tetangga terdekat yang akan menentukan *query* baru masuk ke kelas mana yang ditentukan. Tahap kedua, $K$ tetangga terdekat dicari dengan cara menghitung jarak titik *query* dengan *training* data. Tahap ketiga, setelah mengetahui jarak masing-masing *training* data dengan titik *query*, kemudian lihat nilai yang paling kecil. Tahap keempat, ambil nilai $K$ terkecil selanjutnya lihat kelas yang paling banyak untuk menjadi kelas dari *query* baru (Deza dan Deza, 2009).

Menurut Prasetyo (2014), jauh atau dekatnya jarak titik dengan tetangganya bisa dihitung dengan menggunakan jarak Euclid yang dipresentasikan sebagai berikut:

$$d(x_{ik}, x^*_{jk}) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x^*_{jk})^2} \tag{2.8}$$

dimana:

$d(x_{ik}, x^*_{jk})$ = jarak Euclid data *training* ke-*i* dengan data *testing* ke-*j*

$x_{ik}$ = nilai variabel bebas ke-*k* dari data *training* ke-*i*, $i = 1, 2, ..., n$

$x^*_{jk}$ = nilai variabel bebas ke-*k* dari data *testing* ke-*j*, $j = 1, 2, ..., n^*$

$p$ = banyaknya variabel bebas

Menurut Prasetyo (2012), pada metode K-NN, nilai $K$ menyatakan jumlah tetangga terdekat yang dilibatkan dalam proses penentuan prediksi label kelas pada data *testing*. Dari $K$ tetangga terdekat yang terpilih kemudian dilakukan voting kelas dari nilai $K$ tersebut. Kelas dengan jumlah suara tetangga terbanyaklah yang diberikan sebagai label kelas hasil prediksi pada data *training* tersebut. Berikut adalah algoritma K-NN.

1. Menentukan jumlah *K* tetangga terdekat.
2. Melakukan perhitungan jarak antar data *testing* dan semua data *training* menggunakan rumus jarak Euclid.
3. Mengurutkan jarak (*ranking*).
4. Menggunakan *voting* kelas sebagai prediksi dari data *testing* tersebut.

Ada beberapa hal yang mempengaruhi kinerja K-NN, diantaranya adalah pemilihan nilai $K$. Jika $K$ terlalu kecil maka berakibat hasil prediksi yang didapat bisa sensitif terhadap keberadaan *noise*. Jika $K$ terlalu besar maka tetangga terdekat yang terpilih terlalu banyak dari kelas lain yang sebenarnya tidak relevan karena jarak yang terlalu jauh. Pemilihan nilai $K$ genap atau ganjil juga menjadi perhatian. Untuk $K$ genap dengan jumlah klasifikasi genap akan ada kemungkinan *voting* dari kedua klasifikasi mendapat suara yang sama. Akan tetapi untuk $K$ ganjil dengan jumlah klasifikasi genap akan memudahkan karena dijamin kedua kelas tidak akan mendapat suara yang sama (Prasetyo, 2012).

Ketepatan algoritma K-NN sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur agar performa klasifikasi menjadi lebih baik.

K-NN memiliki beberapa kelebihan yaitu ketangguhan terhadap data *training* yang memiliki banyak *noise* dan efektif apabila data *training*-nya besar. Sedangkan, kelemahannya adalah K-NN perlu menentukan nilai dari parameter $K$ (jumlah dari tetangga terdekat), data *training* berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan dan atribut mana yang harus digunakan untuk mendapatkan hasil terbaik, dan biaya komputasi cukup tinggi

karena diperlukan perhitungan jarak dari tiap titik *query* pada keseluruhan data *training* (Han dan Kamber, 2006).

## 2.3 Data *Training* dan Data *Testing*

Menurut Prasetyo (2014), data untuk pengujian klasifikasi dibagi menjadi data *training* dan data *testing*. Data atau vektor yang sudah diketahui sebelumnya untuk label kelompok dan digunakan untuk membangun model *classifier* disebut data *training*. Data atau vektor yang belum diketahui (dianggap belum diketahui) label kelompoknya sehingga menggunakan model *classifier* yang sudah dibangun untuk diketahui label kelompoknya disebut data *testing*.

Data *training* digunakan oleh algoritma klasifikasi untuk membentuk sebuah model *classifier*. Model ini merupakan representasi pengetahuan yang akan digunakan untuk memprediksi kelompok dari data baru yang belum pernah ada. Data testing digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Data yang ada pada data *testing* seharusnya tidak boleh ada pada data *training* sehingga dapat diketahui apakah model *classifier* sudah tepat atau belum dalam melakukan klasifikasi (Witten dan Eibe, 2011).

Model klasifikasi kemudian dibangun berdasarkan data *training* dan data *testing* dan kemudian kinerjanya diukur berdasarkan data *testing*. Proporsi pembagian data *training* dan data *testing* biasanya diskrit, misalnya 80:20 (artinya 80% sebagai data *training* dan 20% data *testing*) serta 50:50 (artinya 50% data *training* dan 50% data *testing*). Jumlah data *training* dan data *testing* dapat dihitung menggunakan Persamaan (2.9) dan Persamaan (2.10) sebagai berikut.

$$\text{Jumlah data } \textit{training} = \text{ proporsi data } \textit{training} \times N \tag{2.9}$$

$$\text{Jumlah data } \textit{testing} = N - \text{ jumlah data } \textit{training} \tag{2.10}$$

dengan:

$N$ = jumlah seluruh data (data *training* + data *testing*)

Pembagian data *training* dan data *testing* dengan metode proporsi adalah metode yang paling sederhana namun memiliki beberapa keterbatasan. Jumlah data *training* lebih sedikit yang tersedia untuk pelatihan karena sebagian harus

digunakan untuk data *testing*. Akibatnya, model yang dibangun kemungkinan tidak sebaik ketika semua data digunakan sebagai data *training*. Model yang dibangun juga sangat tergantung pada komposisi pemecahan *set* data *training* dan data *testing*. Semakin sedikit data *training*, maka varian dalam model semakin besar (Prasetyo, 2014).

## 2.4 Pengukuran Tingkat Akurasi

Pengukuran tingkat akurasi klasifikasi baik metode *naive* Bayes maupun K-NN dapat dilakukan dengan menghitung kesalahan klasifikasi. Ukuran yang dapat digunakan adalah *Apparent Error Rate* (APER). Nilai APER menyatakan fraksi atau proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi. Semakin banyak kesalahan klasifikasi akan berdampak pada hasil keakurasian metode pengklasifikasian. Metode pengklasifikasian yang baik akan menghasilkan probabilitas kesalahan klasifikasi yang semakin kecil.

Evaluasi fungsi klasifikasi dilakukan dengan cara membuat tabulasi antara *actual group* dan *predicted group* yang diperoleh dari fungsi klasifikasi dari metode *naive* Bayes dan K-NN yang selanjutnya dilakukan perhitungan proporsi pengamatan yang salah klasifikasi. Diharapkan proporsi pengamatan yang salah diklasifikasikan tersebut bisa sekecil mungkin. Dari masing-masing bagian tersebut dihitung nilai APER sehingga diperoleh ketepatan klasifikasinya. Menghitung nilai APER dapat dilakukan melalui tabel klasifikasi sebagai berikut:

**Tabel 2.2** Tabel Klasifikasi

| *Actual Membership* | *Predicted Membership* | | Total |
|---|---|---|---|
| | $\hat{c}_1$ | $\hat{c}_2$ | |
| $c_1$ | $f_{11}$ | $f_{12}$ | $A$ |
| $c_2$ | $f_{21}$ | $f_{22}$ | $B$ |
| Total | $C$ | $D$ | $E$ |

Keterangan:

$f_{11}$ = jumlah obyek dari $c_1$ tepat diklasifikasikan sebagai $\hat{c}_1$

$f_{12}$ = jumlah obyek dari $c_1$ salah diklasifikasikan sebagai $\hat{c}_2$

$f_{21}$ = jumlah obyek dari $c_2$ tepat diklasifikasikan sebagai $\hat{c}_1$

$f_{22}$ = jumlah obyek dari $c_2$ salah diklasifikasikan sebagai $\hat{c}_2$

dengan perhitungan nilai APER sebagai berikut:

$$APER = \frac{\text{jumlah obyek yang salah klasifikasi}}{\text{jumlah prediksi yang dilakukan}} \times 100\% = \frac{f_{12} + f_{21}}{f_{11} + f_{12} + f_{21} + f_{22}} \times 100\% \quad (2.11)$$

(Johnson dan Wichern, 2002).

## 2.5 Pajak Pertambahan Nilai

Menurut Waluyo (2011), Pajak Pertambahan Nilai (PPN) yaitu penggantian pajak penjualan, karena pajak ini tidak bisa memadai dan mencapai sasaran kebutuhan pembangunan masyarakat dan menampung kegiatannya, kegiatan tersebut yaitu pemerataan dalam membebankan pajak, meningkatkan sumber penerimaan negara, dan mendorong produk ekspor. PPN ialah pajak atas konsumsi barang dan jasa yang dikenakan di dalam negeri (didalam daerah pabean).

PPN yang diterapkan di Indonesia adalah PPN Tipe Konsumsi (*Consumption Type VAT*). Dilihat dari sisi perlakuan terhadap barang modal, artinya seluruh biaya yang dikeluarkan untuk perolehan barang modal dapat dikurangi dari dasar pengenaan pajak. Dalam bahasa *indirect subtraction method*, Pajak Masukan (*input tax*) sehingga barang modal dapat dikreditkan dengan Pajak Keluaran (*output tax*).

### 2.5.1 Dasar Hukum Pajak Pertambahan Nilai

Undang-undang yang mengatur pengenaan Pajak Pertambahan Nilai (PPN) dan Pajak Penjualan atas Barang Mewah adalah Undang-Undang Nomor 8 Tahun 1983 tentang Pajak Pertambahan Nilai Barang dan Jasa Kena Pajak Penjualan atas Barang Mewah sebagaimana telah beberapa kali diubah terakhir dengan Undang-Undang Nomor 42 Tahun 2009.

### 2.5.2 Subjek Pajak Pertambahan Nilai

Menurut Resmi (2017) pihak atau subjek yang harus memungut, menyetor, dan melaporkan PPN yaitu :

a. Pengusaha Kena Pajak (PKP) menyerahkan barang kena pajak dan jasa kena pajak di dalam daerah pabean serta melaksanakan pengiriman barang atau jasa (ekspor) yang terkena pajak atau barang tidak berwujud ataupun berwujud.
b. Pengusaha kecil memilih supaya dikukuhkan menjadi PKP, maksudnya pengusaha kecil menyerahkan Barang Kena Pajak (BKP) atau Jasa Kena Pajak (JKP) dan jumlah dari peredaran atau penerimaan brutonya dibawah Rp 600.000.000,- pertahun. Pengusaha tersebut menginginkan supaya dikukuhkan sebagai PKP, dan selanjutnya harus melaksanakan kewajiban layaknya PKP.

### 2.5.3 Objek Pajak Pertambahan Nilai

Objek PPN sesuai dengan diberlakukannya UU yang baru. UU No. 42 Tahun 2009 mengalami perubahan dan berlaku mulai 1 April 2010. Objek PPN dikenakan pada:

a. Pengusaha melakukan penyerahan BKP di dalam Daerah Pabean
b. Impor BKP
c. Pengusaha melakukan Penyerahan JKP di dalam Daerah Pabean
d. Pemanfaatan BKP tidak berwujud baik di dalam maupun di luar Daerah Pabean
e. Pemanfaatan JKP baik didalam maupun luar Daerah Pabean
f. Ekspor BKP Berwujud dari PKP
g. Ekspor BKP Tak Berwujud dari PKP
h. Ekspor JKP oleh PKP

### 2.5.4 Tarif Pajak Pertambahan Nilai

Tarif dasar pengenaan pajak yaitu:

a. Tarif untuk PPN adalah sebesar 10%.
b. Ekspor Barang Kena Pajak tarif PPN sebesar 0%, yang dimaksud 0% tidak mempunyai arti bebas dari pengenaan PPN, akan tetapi pajak

masukan dari barang yang diekspor bisa dikreditkan. Sedangkan berdasarkan Peraturan Pemerintah tarif PPN dapat diubah dengan minimalnya 5% dan maksimalnya 15% dan Indonesia masih tetap menggunakan prinsip tarif tunggal.

# BAB 3
# METODOLOGI PENELITIAN

## 3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada bulan Februari sampai dengan bulan Mei 2020, bertempat di Laboratorium Statistika Komputasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman Samarinda dengan tempat pengambilan data di KPP Pratama Samarinda Ulu.

## 3.2 Rancangan Penelitian

Penelitian yang dilakukan merupakan penelitian *non*-eksperimen, karena dilakukan dengan cara pengambilan data yang sudah tersedia. Penelitian menggunakan rancangan kausal komparatif yang bersifat *ex-post facto*, artinya data dikumpulkan setelah semua kejadian yang dipersoalkan berlangsung.

## 3.3 Populasi dan Sampel Penelitian

Populasi adalah objek atau subjek yang berada pada suatu wilayah dan memenuhi syarat-syarat tertentu berkaitan dengan masalah penelitian, sedangkan sampel adalah sebagian dari populasi yang karakteristiknya hendak diselidiki dan dianggap dapat mewakili keseluruhan populasi (Riduwan dan Akdon, 2010). Penggunaan sampel dilakukan bila populasi besar dan peneliti tidak mungkin mempelajari semua yang ada pada populasi, misalnya karena keterbatasan dana, tenaga, dan waktu maka peneliti dapat menggunakan sampel yang diambil dari populasi itu (Sugiyono, 2010).

Dalam penelitian ini yang menjadi populasi adalah seluruh Wajib Pajak Badan di KPP Pratama Samarinda Ulu dan yang menjadi sampel adalah Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018.

## 3.4 Teknik Sampling

Pada penelitian ini metode yang digunakan dalam penarikan sampel adalah *purposive* sampling. Menurut Sugiyono (2010), *purposive* sampling adalah teknik

penentuan sampel dengan pertimbangan tertentu. Sampling ini dikenal juga sebagai sampling pertimbangan di mana sampel yang digunakan dengan pertimbangan ketersediaan data terbaru di KPP Pratama Samarinda Ulu.

### 3.5 Teknik Pengambilan Data

Data yang digunakan dalam penelitian ini adalah data sekunder. Pengumpulan data sekunder adalah teknik pengumpulan data yang diperoleh dari instansi terkait, dalam hal ini adalah KPP Pratama Samarinda Ulu. Adapun data terkait yang digunakan adalah data Wajib Pajak Badan mengenai status pembayaran pajak, pendapatan, bentuk badan, dan status pelaporan pajak.

### 3.6 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari variabel terikat $(Y)$ dengan dua kategori dan tiga variabel bebas $(X)$ dengan pengelompokan variabelnya berdasarkan pengelompokan pada Kantor Pelayanan Pajak Pratama Samarinda Ulu. Keterangan untuk tiap variabel dapat dilihat pada Tabel 3.1.

**Tabel 3.1** Variabel Penelitian

| **Variabel Penelitian** | **Keterangan** | **Kategori** |
|---|---|---|
| Status Pembayaran PPN $(Y)$ | Status pembayaran PPN dikatakan patuh jika wajib pajak badan melaporkan penghasilannya dan melaksanakan pelunasan atas pajak terutang. | $Y = \begin{cases} 1 \text{ jika patuh membayar PPN} \\ 2 \text{ jika tidak patuh membayar PPN} \end{cases}$ |

**Tabel 3.1** Variabel Penelitian (Lanjutan)

| Variabel Penelitian | Keterangan | Kategori |
|---|---|---|
| Pendapatan $(X_1)$ | Pendapatan wajib pajak badan pertahun pada tahun 2018 (Rupiah). | $X_1 = \begin{cases} 1 \text{ jika pendapatan <100 Juta} \\ 2 \text{ jika pendapatan 100 - 500 Juta} \\ 3 \text{ jika pendapatan >500 Juta - 1 Miliar} \\ 4 \text{ jika pendapatan >1 - 10 Miliar} \\ 5 \text{ jika pendapatan >10 Miliar} \end{cases}$ |
| Bentuk Badan $(X_2)$ | Bentuk badan wajib pajak pada saat mendaftar di KPP Pratama Samarinda Ulu. | $X_2 = \begin{cases} 1 \text{ jika bentuk badan PT} \\ 2 \text{ jika bentuk badan CV} \\ 3 \text{ jika bentuk badan BUMD/BUMN} \\ 4 \text{ jika bentuk badan koperasi} \\ 5 \text{ jika bentuk badan lainnya} \end{cases}$ |
| Status Pelaporan $(X_3)$ | Status pelaporan dikatakan tepat waktu jika wajib pajak melaporkan SPT Masa PPN dalam jangka waktu yang ditentukan, dan dikatakan tidak tepat waktu jika di luar jangka waktu yang ditentukan | $X_3 = \begin{cases} 1 \text{ jika pelaporan tepat waktu} \\ 2 \text{ jika pelaporan tidak tepat waktu} \end{cases}$ |

## 3.7 Teknik Analisis Data

Teknik analisis data pada penelitian ini yaitu mengklasifikasikan status pembayaran pajak pertambahan nilai menggunakan metode *naive* Bayes dan K-NN. Penelitian ini menggunakan bantuan *software* komputer yaitu R. Adapun

langkah-langkah dalam teknik analisis data yang dilakukan peneliti adalah sebagai berikut:

1. Analisis Statistika Deskriptif

   Analisis statistika deskriptif adalah suatu metode yang digunakan untuk menggambarkan atau mengkarakteristikkan suatu obyek penelitian. Penyajian data dilakukan dengan membuat tabulasi penyajian data dalam bentuk diagram *pie*. Tujuan analisis ini adalah agar data yang dipaparkan mudah dipahami dan dimengerti. Pada analisis ini menggunakan bantuan *software* R untuk menyajikan data dalam bentuk diagram *pie* dari data Wajib Pajak Badan berdasarkan variabel pendapatan, bentuk badan dan status pelaporan pajak.

2. Randomisasi Data

   Dilakukan randomisasi data dengan menggunakan bantuan *software* R. Tujuan dilakukannya randomisasi data adalah agar semua data memiliki kesempatan yang sama untuk menjadi data *training* dan data *testing*.

3. Menentukan Data *Training* dan Data *Testing*

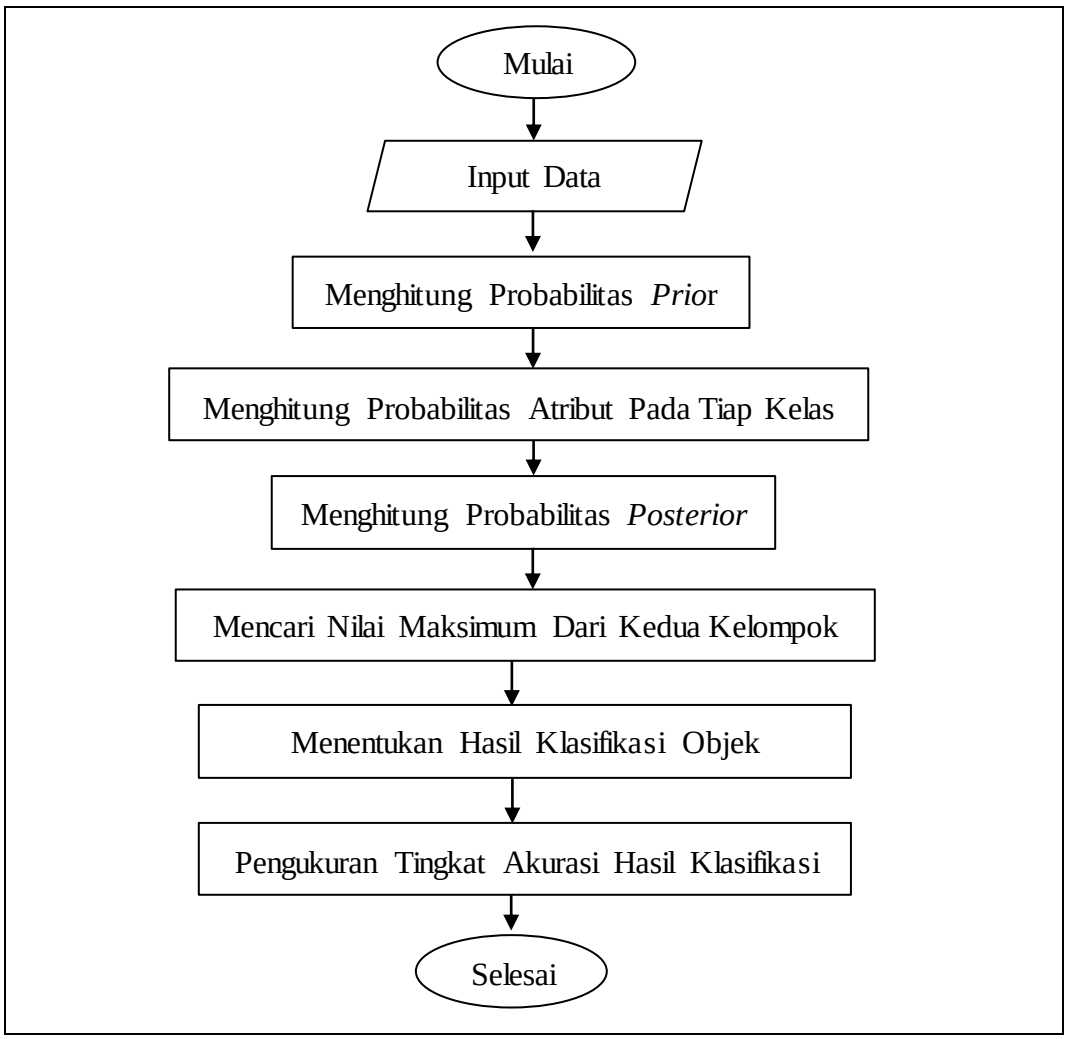
   Untuk menentukan data *training* dan data *testing*, data dibagi menjadi dua dengan cara memilih 80% dari data urutan teratas menjadi data *training*, dan sisanya sebanyak 20% urutan terbawah sebagai data *testing*.

4. Metode *naive* Bayes

   Langkah-langkah yang dilakukan dalam mengolah data pengklasifikasian dengan metode *naive* Bayes menggunakan bantuan *software* R adalah sebagai berikut:

   a. Menghitung nilai probabilitas awal (*prior*) berdasarkan Persamaan (2.4).
   b. Menghitung probabilitas setiap variabel bebas pada setiap kelompok berdasarkan Persamaan (2.6).
   c. Menghitung probabilitas akhir (*posterior*) pada masing-masing kelas berdasarkan persamaan (2.7).
   d. Mencari nilai maksimum pada kedua kelompok dan menentukan hasil klasifikasi obyek.
   e. Mengevaluasi hasil klasifikasi dengan menghitung nilai APER.

Tahapan metode *naive* Bayes lebih jelasnya dapat dilihat pada Gambar 3.1 berikut ini :




**Gambar 3.1** Diagram Alir Metode *Naive* Bayes

5. Metode K-*Nearest Neighbor* (K-NN)

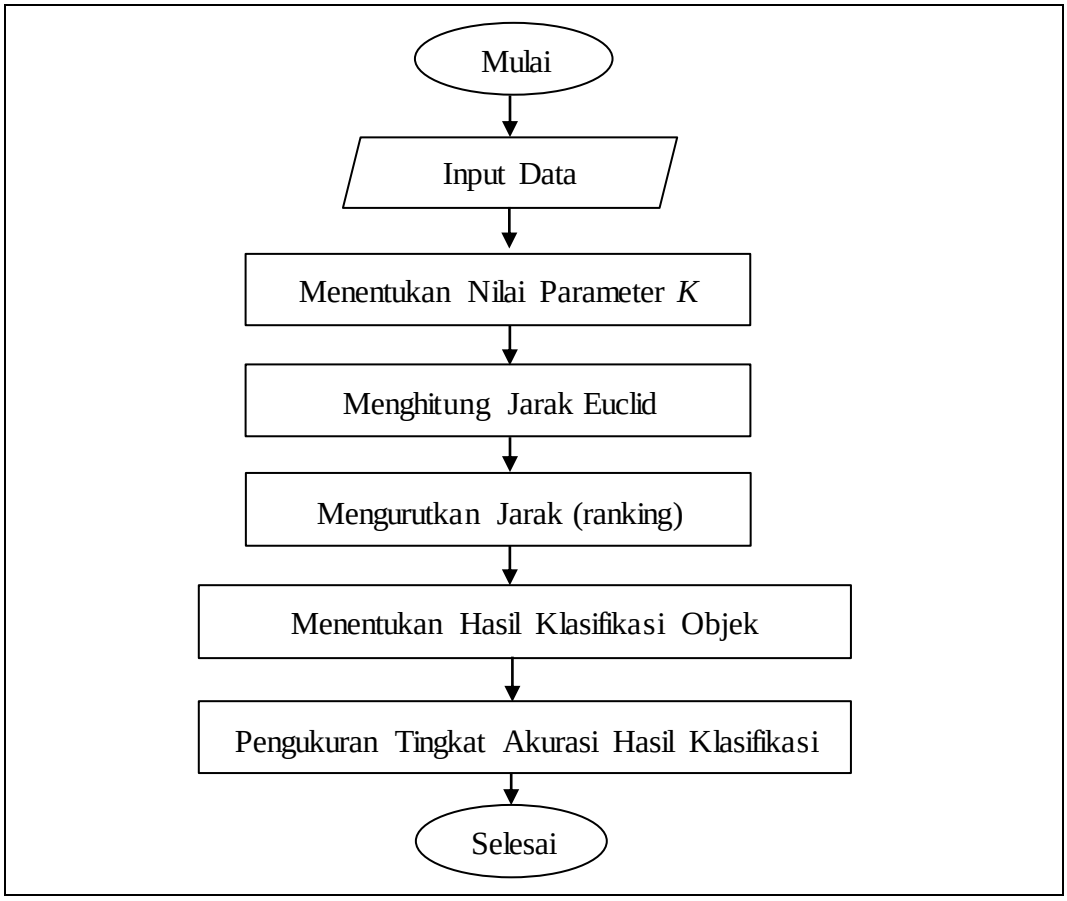
   Langkah-langkah yang dilakukan dalam mengolah data pengklasifikasian dengan metode K-NN adalah sebagai berikut:

   a. Menentukan nilai parameter $K$, pada penelitian ini menggunakan $K = 3$, $K = 5$, dan $K = 7$.
   b. Menghitung jarak Euclid objek terhadap data *training* yang diberikan berdasarkan Persamaan (2.8)
   c. Mengurutkan data yang mempunyai jarak terkecil hingga terbesar.

d. Menentukan hasil klasifikasi dengan menggunakan kategori *nearest neighbor* yang paling banyak.

e. Mengevaluasi hasil klasifikasi dengan menghitung nilai APER.

Tahapan metode *naive* Bayes lebih jelasnya dapat dilihat pada Gambar 3.2 berikut ini :



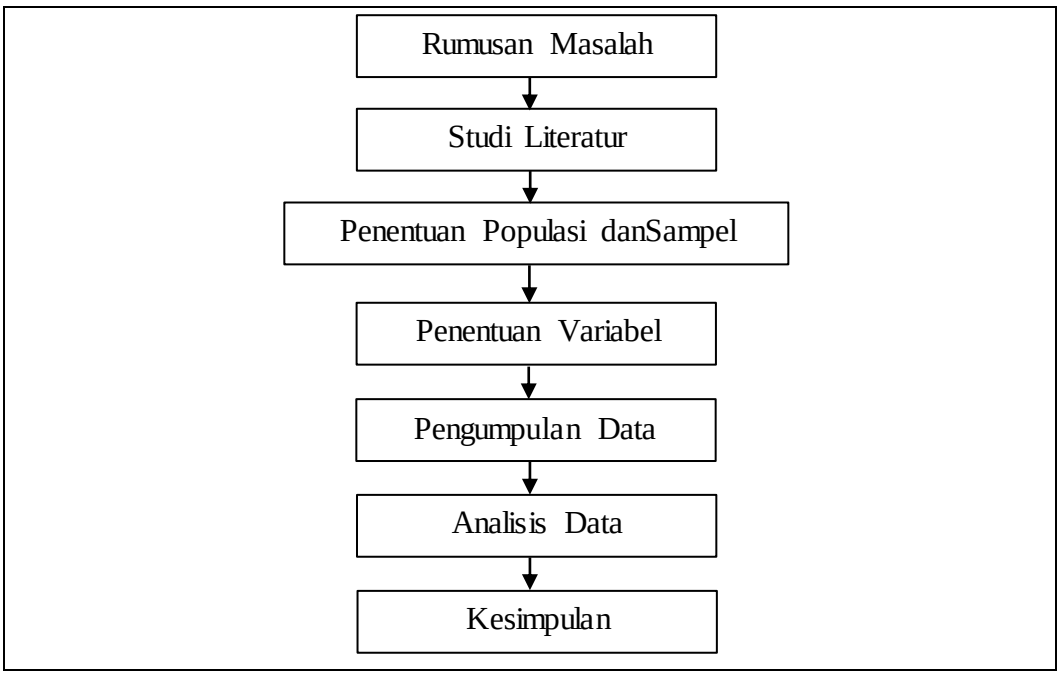

**Gambar 3.2** Diagram Alir Metode K-NN

6. Membandingkan Hasil Klasifikasi

Setelah diperoleh hasil pengukuran tingkat akurasi menggunakan APER baik pada metode *naive* Bayes maupun metode K-NN, selanjutnya dilakukan perbandingan hasil klasifikasi. Ketepatan klasifikasi dengan nilai yang terbesar akan diambil untuk kemudian ditarik kesimpulan.

## 3.8 Kerangka Penelitian

Secara umum penelitian ini memiliki kerangka kerja seperti berikut :




**Gambar 3.3** Kerangka Penelitian

# BAB 4
# HASIL DAN PEMBAHASAN

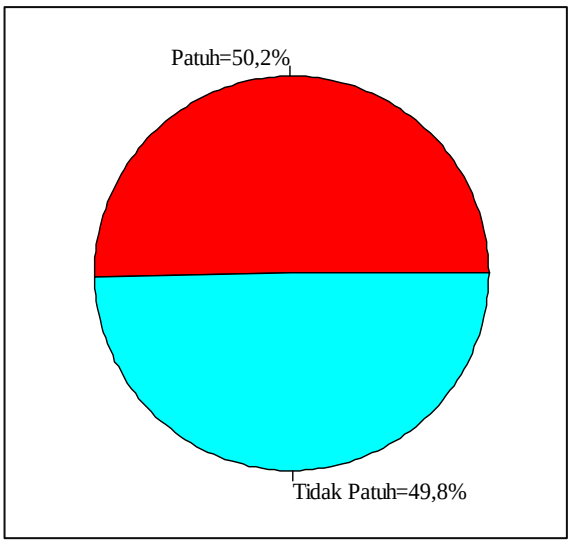
Data yang digunakan dalam penelitian ini merupakan data wajib pajak badan di KPP Pratama Samarinda Ulu Tahun 2018 yang dapat dilihat pada Lampiran 1. Data tersebut akan dianalisis menggunakan metode klasifikasi *naive* Bayes dan K-NN. Adapun variabel terikat yang digunakan adalah status pembayaran PPN pada wajib pajak badan dengan kategori patuh dan tidak patuh dengan variabel bebas adalah pendapatan, bentuk badan dan status pelaporan pajak.

## 4.1 Analisis Statistika Deskriptif

Analisis statistika deskriptif ini dilakukan untuk mengetahui gambaran umum data wajib pajak badan di KPP Pratama Samarinda Ulu Tahun 2018. Karakteristik yang digambarkan pada analisis statistika deskriptif adalah status pembayaran PPN, pendapatan, bentuk badan dan status pelaporan.

1. Persentase data wajib pajak berdasarkan status pembayaran PPN

   Gambaran umum pada analisis statistika deskriptif mengenai status pembayaran PPN pada wajib pajak badan dengan status pembayaran patuh dan tidak patuh dapat dilihat pada Gambar 4.1.

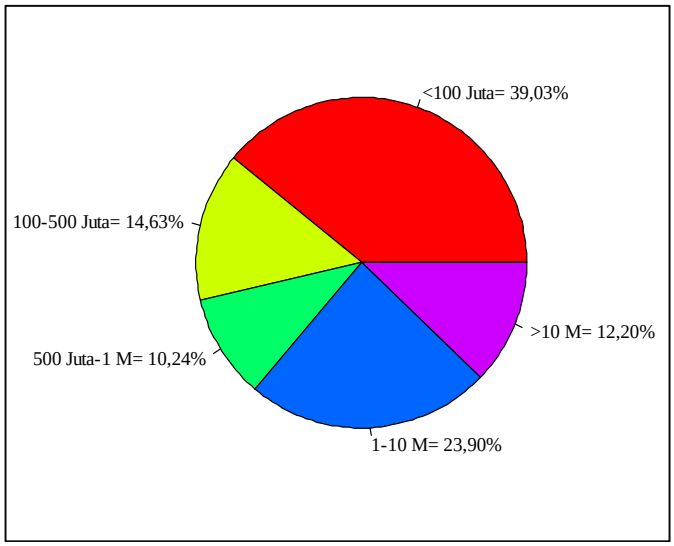



**Gambar 4.1** Persentase status pembayaran pajak

Pada Gambar 4.1 menunjukkan persentase status pembayaran pajak pertambahan nilai di KPP Pratama Samarinda Ulu Tahun 2018. Dari 205 data wajib pajak badan dapat diketahui bahwa 50,2% atau sebanyak 103 wajib pajak badan berstatus patuh membayar PPN dan 49,8% atau sebanyak 102 lainnya merupakan wajib pajak badan dengan status tidak patuh membayar PPN. Dari data tersebut dapat diketahui bahwa wajib pajak badan dengan status patuh memiliki persentase yang lebih besar dibandingkan dengan wajib pajak badan yang berstatus tidak patuh membayar PPN.

2. Persentase data wajib pajak berdasarkan pendapatan

Persentase wajib pajak badan di KPP Pratama Samarinda Ulu berdasarkan pendapatan pertahun dapat dilihat pada Gambar 4.2.



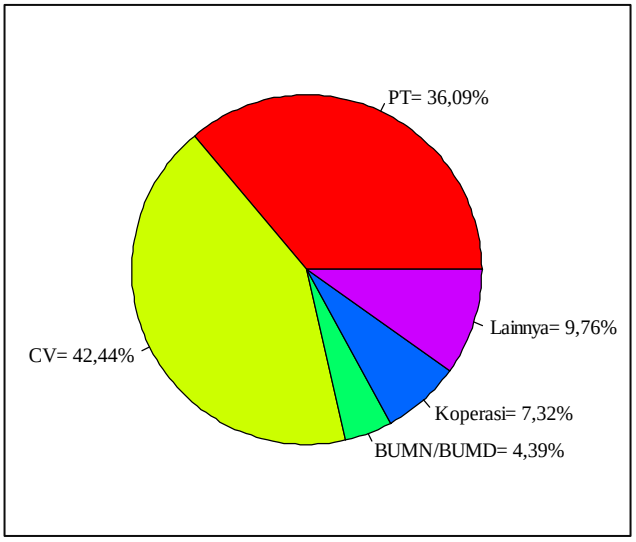

**Gambar 4.2** Persentase pendapatan wajib pajak badan

Berdasarkan Gambar 4.2 dapat diketahui bahwa wajib pajak badan dengan pendapatan sebesar < 100 Juta berjumlah 80 wajib pajak (39,03%). Wajib pajak badan dengan pendapatan sebesar 100-500 Juta berjumlah 30 wajib pajak (14,63%). Wajib pajak badan dengan pendapatan sebesar 500 Juta – 1 M berjumlah 21 wajib pajak (10,24%). Wajib pajak badan dengan pendapatan sebesar 1-10 M berjumlah 49 wajib pajak (23,90%). Wajib pajak badan dengan pendapatan sebesar > 10 M berjumlah 25 wajib pajak (12,20%). Dari data tersebut dapat disimpulkan bahwa wajib pajak badan di KPP Pratama

Samarinda Ulu dengan pendapatan sebesar < 100 Juta memiliki jumlah terbanyak dengan persentase sebesar 39,03%.

3. Persentase data wajib pajak berdasarkan bentuk badan

   Persentase wajib pajak badan di KPP Pratama Samarinda Ulu berdasarkan bentuk badan dapat dilihat pada Gambar 4.3.

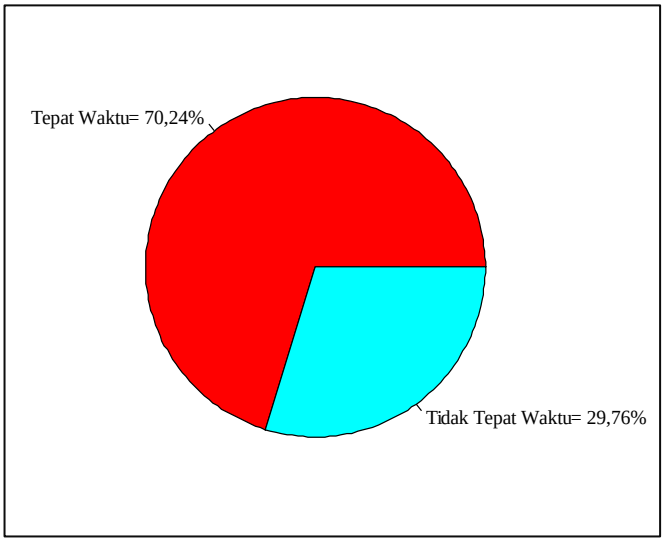



**Gambar 4.3** Persentase bentuk badan

Berdasarkan Gambar 4.3 dapat diketahui bahwa dari 205 data wajib pajak badan di KPP Pratama Samarinda Ulu, sebanyak 74 wajib pajak (36,09%) berbentuk PT, sebanyak 87 wajib pajak (42,44%) berbentuk CV, sebanyak 9 wajib pajak (4,39%) berbentuk BUMN/BUMD, sebanyak 15 wajib pajak (7,32%) berbentuk Koperasi dan sebanyak 20 wajib pajak (9,76%) merupakan organisasi lainnya seperti Firma, Ormas dan Yayasan. Dari data tersebut dapat disimpulkan bahwa wajib pajak badan di KPP Pratama Samarinda Ulu paling banyak berbentuk CV.

4. Persentase data wajib pajak berdasarkan status pelaporan pajak

   Persentase wajib pajak badan di KPP Pratama Samarinda Ulu berdasarkan status pelaporan pajak dapat dilihat pada Gambar 4.4.




**Gambar 4.4** Persentase status pelaporan pajak

Berdasarkan Gambar 4.4 dapat diketahui bahwa wajib pajak badan di KPP Pratama Samarinda Ulu lebih banyak berstatus tepat waktu dalam pelaporan pajak yaitu berjumlah 144 wajib pajak (70,24%) dibandingkan dengan status tidak tepat waktu dalam pelaporan pajak yang berjumlah 61 wajib pajak (29,76%).

## 4.2 Membagi Data *Training* dan Data *Testing*

Proses pertama sebelum melakukan analisis klasifikasi data wajib pajak badan dengan menggunakan metode *naive* Bayes dan K-NN, data data wajib pajak badan dibagi menjadi data *training* dan data *testing* dengan proporsi pembagian 80:20. Adapun perhitungan pembagian data training dan data testing dengan menggunakan Persamaan (2.9) dan Persamaan (2.10) sebagai berikut:

$$\begin{aligned}\text{Jumlah data } \textit{training} &= \text{Proporsi data } \textit{training} \times N \\ &= 80\% \times 205 \\ &= 164\end{aligned}$$

$$\begin{aligned}\text{Jumlah data } \textit{testing} &= N - \text{Jumlah data } \textit{training} \\ &= 205 - 164 \\ &= 41\end{aligned}$$

Berdasarkan perhitungan diperoleh hasil bahwa 164 data pertama hasil randomisasi data untuk masing-masing variabel penelitian berfungsi sebagai data

*training* yang dapat dilihat pada Lampiran 5. Data *training* ini akan digunakan dalam proses perhitungan pada metode *naive* Bayes serta perhitungan jarak pada metode K-NN, sedangkan 41 data terakhir untuk masing-masing variabel penelitian berfungsi sebagai data *testing* yang nantinya digunakan untuk melakukan proses klasifisikasi dan pengukuran tingkat akurasi. Untuk mendapatkan akurasi terbaik dilakukan 5 kali randomisasi data dengan proporsi data *training* dan data *testing* yang sama yaitu 80:20.

## 4.3 Klasifikasi *naive* Bayes

Pada proses klasifikasi metode *naive* Bayes terdapat tiga alur yaitu menghitung probabilitas *prior*, menghitung probabilitas setiap variabel bebas pada setiap kelompok dan menghitung perkalian probabilitas *prior* dan probabilitas variabel bebas pada setiap kelompok. Adapun data yang digunakan untuk perhitungan klasifikasi adalah menggunakan data *training* 80% sebanyak 164 data, sedangkan data *testing* 20% sebanyak 41 data. Pada kali ini akan diberikan contoh perhitungan menggunakan data randomisasi ke 4. Adapun tahapan-tahapan klasifikasi metode *naive* Bayes dalam pengklasifikasian status pembayaran pajak sebagai berikut:

1. Menghitung probabilitas *prior* (awal) setiap kelompok

   Tahapan klasifikasi metode *naive* Bayes yang pertama yaitu menghitung probabilitas *prior* pada kedua kelompok menggunakan data *training*. Adapun nilai probabilitas *prior* setiap kelompok menggunakan Persamaan (2.4) sebagai berikut:

   a. Kelompok pertama (wajib pajak badan dengan status patuh membayar PPN)

      Pada 164 data *training* yang ada, terdapat 83 wajib pajak badan yang berstatus patuh membayar pajak, sehingga nilai probabilitas *prior* kelompok pertama sebagai berikut:

$$P(Y_1) = \frac{83}{164}$$

$$= 0,5061$$

Jadi, probabilitas *prior* wajib pajak badan dengan status patuh membayar PPN adalah sebesar 0,5061.

b. Kelompok kedua (wajib pajak badan dengan status tidak patuh membayar PPN)

Pada 164 data *training* yang ada, terdapat 81 wajib pajak badan dengan status tidak patuh membayar pajak, sehingga nilai probabilitas *prior* kelompok kedua sebagai berikut:

$$P(Y_2) = \frac{81}{164}$$

$$= 0,4939$$

Jadi, probabilitas *prior* wajib pajak badan dengan status tidak patuh membayar PPN adalah sebesar 0,4939. Adapun nilai probabilitas *prior* wajib pajak badan pada setiap kelompok dapat dilihat pada Tabel 4.1.

**Tabel 4.1** Probabilitas *Prior* Pada Setiap Kelompok

| **Status Pembayaran PPN** | **Jumlah** | **Probabilitas** |
|---|---|---|
| Patuh | 83 | 0,5061 |
| Tidak Patuh | 81 | 0,4939 |

2. Menghitung probabilitas setiap variabel bebas pada setiap kelompok

Perhitungan nilai probabilitas setiap variabel bebas pada kedua kelompok dilakukan pada data *testing* berdasarkan data *training*. Pada data *testing* pertama diketahui wajib pajak badan memiliki pendapatan $(X_1)$ sebesar 1-10 Miliar Rupiah, bentuk badan $(X_2)$ berupa PT dan memiliki status pelaporan pajak $(X_3)$ tepat waktu, sehingga perhitungan nilai probabilitas setiap variabel bebas pada kedua kelompok menggunakan Persamaan (2.6) sebagai berikut:

a. Pendapatan $(X_1)$

Adapun nilai probabilitas variabel pendapatan data *testing* pertama pada setiap kelompok sebagai berikut :

Probabilitas variabel pendapatan untuk status patuh membayar pajak dihitung sebagai berikut:

$$P(X_1\text{=1-10 Miliar} \mid \text{Patuh}) = \frac{19}{83}$$
$$= 0,2289$$

Sehingga probabilitas variabel pendapatan sebesar 1-10 Miliar Rupiah dengan syarat status patuh membayar pajak adalah sebesar 0,2289. Probabilitas variabel pendapatan untuk status tidak patuh membayar pajak dihitung sebagai berikut:

$$P(X_1\text{=1-10 Miliar} \mid \text{TidakPatuh}) = \frac{20}{81}$$
$$= 0,2469$$

Sehingga probabilitas variabel pendapatan sebesar 1-10 Miliar Rupiah dengan syarat status tidak patuh membayar pajak adalah sebesar 0,2469. Adapun nilai probabilitas variabel pendapatan pada setiap kelompok dapat dilihat pada Tabel 4.2. Angka di dalam kurung menyatakan nilai probabilitas variabel pendapatan.

**Tabel 4.2** Probabilitas Variabel Pendapatan Pada Setiap Kelompok

| **Pendapatan** | **Status Pembayaran Pajak** | |
|---|---|---|
| | **Patuh** | **Tidak Patuh** |
| < 100 Juta | 35 (0,4217) | 30 (0,3704) |
| 100 – 500 Juta | 8 (0,0964) | 16 (0,1975) |
| 500 Juta – 1 Miliar | 9 (0,1084) | 8 (0,0988) |
| 1 – 10 Miliar | 19 (0,2289) | 20 (0,2469) |
| > 10 Miliar | 12 (0,1446) | 7 (0,0864) |

Sumber: Lampiran 7

b. Bentuk Badan $(X_2)$

Adapun nilai probabilitas variabel bentuk badan data *testing* pertama pada setiap kelompok sebagai berikut:

Probabilitas variabel bentuk badan untuk status patuh membayar pajak dihitung sebagai berikut:

$$P(X_2 = \text{PT} \mid \text{Patuh}) = \frac{36}{83}$$
$$= 0,4337$$

Sehingga probabilitas variabel bentuk badan berupa PT dengan syarat status patuh membayar pajak adalah sebesar 0,4337. Probabilitas variabel bentuk badan untuk status tidak patuh membayar pajak dihitung sebagai berikut:

$$P(X_2 = \text{PT} \mid \text{Tidak Patuh}) = \frac{27}{81}$$
$$= 0,3333$$

Sehingga probabilitas variabel bentuk badan berupa PT dengan syarat status tidak patuh membayar pajak adalah sebesar 0,3333. Adapun nilai probabilitas variabel bentuk badan pada setiap kelompok dapat dilihat pada Tabel 4.3. Angka di dalam kurung menyatakan nilai probabilitas variabel bentuk badan.

**Tabel 4.3** Probabilitas Variabel Bentuk Badan Pada Setiap Kelompok

| **Bentuk Badan** | **Status Pembayaran Pajak** | |
|---|---|---|
| | **Patuh** | **Tidak Patuh** |
| PT | 36 (0,4337) | 27 (0,3333) |
| CV | 36 (0,4337) | 30 (0,3704) |
| BUMN/BUMD | 0 (0,0000) | 7 (0,0864) |
| Koperasi | 3 (0,0361) | 6 (0,0741) |
| Lainnya | 8 (0,0964) | 11 (0,1358) |

Sumber: Lampiran 7

c. Status Pelaporan $(X_3)$

Adapun nilai probabilitas variabel status pelaporan data *testing* pertama pada setiap kelompok sebagai berikut:

Probabilitas variabel status pelaporan untuk status patuh membayar pajak dihitung sebagai berikut:

$$P(X_3 = \text{Tepat Waktu} \mid \text{Patuh}) = \frac{65}{83}$$
$$= 0,7831$$

Sehingga probabilitas variabel status pelaporan tepat waktu dengan syarat status patuh membayar pajak adalah sebesar 0,7831. Probabilitas variabel status pelaporan untuk status tidak patuh membayar pajak dihitung sebagai berikut:

$$P(X_3 = \text{Tepat Waktu} \mid \text{Tidak Patuh}) = \frac{53}{81}$$
$$= 0,6543$$

Sehingga probabilitas variabel status pelaporan tepat waktu dengan syarat status tidak patuh membayar pajak adalah sebesar 0,6543. Adapun nilai probabilitas variabel status pelaporan pada setiap kelompok dapat dilihat pada Tabel 4.4. Angka di dalam kurung menyatakan nilai probabilitas variabel status pelaporan.

**Tabel 4.4** Probabilitas Variabel Status Pelaporan Pada Setiap Kelompok

| **Status Pelaporan** | **Status Pembayaran Pajak** | |
|---|---|---|
| | **Patuh** | **Tidak Patuh** |
| Tepat Waktu | 65 (0,7831) | 53 (0,6543) |
| Tidak Tepat Waktu | 18 (0,2169) | 28 (0,3457) |

Sumber: Lampiran 7

3. Menghitung perkalian probabilitas *prior* dan probabilitas setiap variabel bebas pada setiap kelompok (*posterior*)

   Tahap ketiga adalah menghitung perkalian probabilitas *prior* dan probabilitas *posterior* pada data *testing*. Pada data *testing* pertama diketahui wajib pajak badan memiliki pendapatan $(X_1)$ sebesar 1-10 Miliar Rupiah, bentuk badan $(X_2)$ berupa PT dan memiliki status pelaporan pajak $(X_3)$ tepat waktu. Adapun perhitungan perkalian probabilitas *prior* dan probabilitas *posterior* menggunakan Persamaan (2.7) sebagai berikut:

   a. Kelompok pertama (wajib pajak badan dengan status patuh membayar pajak)

$$\begin{aligned} P(Y_1 \mid X_1, X_2, X_3) &= P(Y_1) \times P(X_1 \mid Y_1) \times P(X_2 \mid Y_1) \times P(X_3 \mid Y_1) \\ &= P(\text{Patuh}) \times P(X_1 = \text{1-10 Miliar} \mid \text{Patuh}) \times P(X_2 = \text{PT} \mid \\ &\quad \text{Patuh}) \times P(X_3 = \text{Tepat Waktu} \mid \text{Patuh}) \\ &= 0,5061 \times 0,2289 \times 0,4337 \times 0,7831 \\ &= 0,0393 \end{aligned}$$

   Sehingga dapat diketahui bahwa probabilitas wajib pajak yang berstatus patuh membayar pajak dengan syarat pendapatan sebesar 1-10 Miliar Rupiah, bentuk badan berupa PT dan status pelaporan tepat waktu adalah sebesar 0,0393.

   b. Kelompok kedua (wajib pajak badan dengan status tidak patuh membayar pajak)

$$\begin{aligned} P(Y_2 \mid X_1, X_2, X_3) &= P(Y_2) \times P(X_1 \mid Y_2) \times P(X_2 \mid Y_2) \times P(X_3 \mid Y_2) \\ &= P(\text{Tidak Patuh}) \times P(X_1 = 1-10 \text{ Miliar} \mid \text{Tidak Patuh}) \\ &\quad \times P(X_2 = \text{PT} \mid \text{Tidak Patuh}) \times P(X_3 = \text{Tepat Waktu} \mid \\ &\quad \text{Tidak Patuh}) \\ &= 0,4939 \times 0,2469 \times 0,3333 \times 0,6543 \\ &= 0,0266 \end{aligned}$$

   Sehingga dapat diketahui bahwa probabilitas wajib pajak yang berstatus tidak patuh membayar pajak dengan syarat pendapatan sebesar 1-10 Miliar

Rupiah, bentuk badan berupa PT dan status pelaporan tepat waktu adalah sebesar 0,0266.

Setelah melakukan perhitungan probabilitas *posterior* pada kedua kelompok, dicari nilai maksimum pada kedua kelompok untuk menentukan hasil klasifikasi obyek. Berdasarkan hasil perhitungan dapat diketahui bahwa kelompok yang memiliki *posterior* terbesar adalah wajib pajak dengan kelompok status patuh membayar pajak, yaitu sebesar 0,0393 dibandingkan dengan *posterior* wajib pajak dengan kelompok status tidak patuh membayar pajak, yaitu sebesar 0,0266, sehingga dapat disimpulkan data *testing* pertama yaitu wajib pajak badan dengan pendapatan sebesar 1-10 Miliar Rupiah, bentuk badan berupa PT dan memiliki status pelaporan pajak tepat waktu diklasifikasikan masuk ke dalam kelompok pertama yaitu wajib pajak dengan status patuh membayar pajak. Nilai probabilitas *posterior* pada kedua kelompok untuk data *testing* lainnya dapat dilihat pada Lampiran 8.

4. Mengevaluasi Hasil Klasifikasi

Untuk mengevaluasi hasil klasifikasi pada metode *naive* Bayes dilakukan pengukuran tingkat akurasi dengan menghitung peluang kesalahan klasifikasi menggunakan APER. Pada proses klasifikasi, semakin kecil kesalahan klasifikasi menunjukkan bahwa semakin baik hasil klasifikasi yang didapatkan. Pada hasil klasifikasi, jumlah obyek yang tepat dan salah diklasifikasikan untuk masing-masing kelompok dapat dilihat pada Tabel 4.5. Tanda (*) pada angka-angka menyatakan jumlah obyek kelompok tertentu yang salah diklasifikasikan dengan menggunakan metode *naive* Bayes.

**Tabel 4.5** Hasil Klasifikasi Metode *Naive* Bayes

| **Klasifikasi Awal Status Pembayaran Pajak** | **Prediksi Klasifikasi Metode *Naive* Bayes** | | **Total** |
|---|---|---|---|
| | Patuh | Tidak Patuh | |
| Patuh | 15 | 5* | 20 |
| Tidak Patuh | 2* | 19 | 21 |
| Total | 17 | 24 | 41 |

Sumber: Lampiran 7

Berdasarkan Tabel 4.5 dapat diketahui bahwa status pembayaran pajak pertambahan nilai di KPP Pratama Samarinda Ulu menggunakan metode *naive* Bayes diperoleh hasil yaitu dari 20 wajib pajak yang memiliki status patuh membayar PPN, terdapat 15 wajib pajak tepat diklasifikasikan memiliki status patuh membayar PPN dan 5 lainnya tidak tepat diklasifikasikan. Sedangkan dari 21 wajib pajak yang memiliki status tidak patuh membayar PPN, terdapat 19 wajib pajak tepat diklasifikasikan memiliki status tidak patuh membayar PPN dan 2 lainnya tidak tepat diklasifikasikan, sehingga berdasarkan Persamaan (2.11) maka diperoleh nilai APER sebagai berikut:

$$\begin{aligned} \text{APER} &= \frac{5+2}{41} \times 100\% \\ &= \frac{7}{41} \times 100\% \\ &= 17,07\% \end{aligned}$$

Jadi, nilai APER (laju *error*) pada hasil klasifikasi *naive* Bayes menggunakan data randomisasi ke-4 adalah sebesar 17,07%. Adapun nilai APER pada hasil klasifikasi menggunakan metode *naive* Bayes untuk semua percobaan randomisasi data disajikan pada Tabel 4.6 berikut:

**Tabel 4.6** Nilai APER Pada Metode *Naive* Bayes untuk Setiap Percobaan Randomisasi Data

| **Randomisasi ke-** | **APER** |
|---|---|
| 1 | 31,71% |
| 2 | 29,27% |
| 3 | 21,95% |
| 4 | 17,07% |
| 5 | 19,51% |

Sumber: Lampiran 7

Berdasarkan Tabel 4.6 dapat diketahui bahwa pada data randomisasi ke-4 mendapatkan nilai APER yang paling kecil, sehingga dapat disimpulkan dari semua percobaan randomisasi data, data randomisasi ke-4 dapat menghasilkan

ketepatan klasifikasi metode *naive* Bayes yang paling akurat. Selanjutnya data randomisasi ke-4 akan dicobakan pada metode K-NN untuk menentukan nilai *K* optimal.

## 4.4 Klasifikasi *K-Nearest Neighbor*

Pada proses klasifikasi metode K-NN terdapat tiga alur yaitu menentukan nilai parameter *K*, menghitung jarak Euclid antara data *training* dan data *testing*, kemudian menentukan *ranking* dari hasil perhitungan jarak. Data yang digunakan untuk perhitungan klasifikasi adalah menggunakan data *training* 80% sebanyak 164 data, sedangkan data *testing* 20% sebanyak 41 data. Adapun tahapan-tahapan dalam proses pengklasifikasian dengan metode K-NN sebagai berikut:

1. Menentukan Nilai Parameter *K*

Klasifikasi status pembayaran PPN menggunakan metode K-NN dilakukan dengan menentukan nilai parameter *K* terlebih dahulu. Adapun nilai *K* yang digunakan yaitu *K*=1, *K*=3, *K*=5, *K*=7 dan *K*=9. Pada kali ini akan diberikan contoh perhitungan menggunakan parameter *K*=1.

2. Menghitung Jarak Euclid

Pada tahap selanjutnya dilakukan perhitungan jarak Euclid pada objek antara data *testing* dan data *training* berdasarkan Persamaan (2.8). Perhitungan jarak pada penelitian ini menggunakan angka yang berasal dari kodingan kategori pada variabel bebas. Adapun perhitungan jarak Euclid antara data *testing* pertama dan data *training* pertama hingga data *training* ke-164 sebagai berikut:

$$\begin{aligned} d(x_1, x_1^*) &= \sqrt{(x_{11} - x_{11}^*)^2 + (x_{12} - x_{12}^*)^2 + (x_{13} - x_{13}^*)^2} \\ &= \sqrt{(1-4)^2 + (2-1)^2 + (1-1)^2} \\ &= 3{,}1623 \end{aligned}$$

$$\begin{aligned} d(x_2, x_1^*) &= \sqrt{(x_{21} - x_{11}^*)^2 + (x_{22} - x_{12}^*)^2 + (x_{23} - x_{13}^*)^2} \\ &= \sqrt{(4-4)^2 + (2-1)^2 + (2-1)^2} \\ &= 1{,}4142 \end{aligned}$$

$$d(x_3, x_1^*) = \sqrt{(x_{31} - x_{11}^*)^2 + (x_{32} - x_{12}^*)^2 + (x_{33} - x_{13}^*)^2}$$

$$= \sqrt{(1-4)^2 + (3-1)^2 + (1-1)^2}$$
$$= 3,6055$$
$$\vdots$$
$$d(x_{164}, x_1^*) = \sqrt{(x_{1641} - x_{11}^*)^2 + (x_{1642} - x_{12}^*)^2 + (x_{1643} - x_{13}^*)^2}$$
$$= \sqrt{(1-4)^2 + (1-1)^2 + (1-1)^2}$$
$$= 3$$

Adapun hasil perhitungan jarak Euclid secara lengkap dapat dilihat pada Lampiran 10.

3. Menentukan *ranking* dari hasil perhitungan jarak

Setelah melakukan perhitungan jarak menggunakan Persamaan (2.8), kemudian dilanjutkan dengan mengurutkan (*ranking*) hasil perhitungan jarak untuk menentukan urutan terkecil hingga terbesar. Adapun *rankinng* jarak *eucliedan* dapat dilihat pada Tabel 4.7.

**Tabel 4.7** Ranking Jarak Euclid Antara Data *Testing* Pertama dan Data *Training*

| ***Rank*** | ***Data Training*** | | **Data *Testing* Pertama** | **Batas K-NN** |
|---|---|---|---|---|
| | **Sampel** | **Klasifikasi** | **d** | |
| 1 | 178 | Tidak Patuh | 0 | 1-NN |
| 2 | 174 | Tidak Patuh | 0 | - |
| 3 | 53 | Patuh | 0 | - |
| 4 | 50 | Patuh | 0 | - |
| 5 | 52 | Patuh | 0 | - |
| 6 | 176 | Tidak Patuh | 0 | - |
| 7 | 55 | Patuh | 0 | - |
| 8 | 173 | Tidak Patuh | 0 | - |
| 9 | 175 | Tidak Patuh | 0 | - |
| 10 | 179 | Tidak Patuh | 0 | - |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

| | | | | |
|---|---|---|---|---|
| 164 | 138 | Tidak Patuh | 5,099 | - |

Berdasarkan Tabel 4.7 dapat diketahui bahwa dengan menggunakan 1-NN, pada data *testing* pertama yaitu wajib pajak badan yang memiliki pendapatan $(X_1)$ sebesar 1-10 Miliar Rupiah, bentuk badan $(X_2)$ berupa PT dan memiliki status pelaporan pajak $(X_3)$ tepat waktu, diprediksi berstatus tidak patuh membayar PPN. Adapun hasil prediksi untuk data *testing* lainnya terdapat pada Lampiran 11.

4. Mengevaluasi Hasil Klasifikasi

Setelah mendapatkan hasil prediksi dari seluruh data testing menggunakan $K$=1, dapat dilakukan evaluasi hasil klasifikasi metode K-NN. Pada proses klasifikasi menggunakan metode K-NN, jumlah obyek yang tepat dan salah diklasifikasikan untuk masing-masing kelompok dapat dilihat pada Tabel 4.8. Tanda (*) pada angka-angka menyatakan jumlah obyek kelompok tertentu yang salah diklasifikasikan dengan menggunakan metode K-NN.

**Tabel 4.8** Hasil Klasifikasi Metode K-NN

| **Klasifikasi Awal Status Pembayaran Pajak** | **Prediksi Klasifikasi Metode K-NN** | | **Total** |
|---|---|---|---|
| | Patuh | Tidak Patuh | |
| Patuh | 15 | 5* | 20 |
| Tidak Patuh | 3* | 18 | 21 |
| Total | 18 | 23 | 41 |

Sumber: Lampiran 9

Berdasarkan Tabel 4.8 dapat diketahui bahwa status pembayaran pajak pertambahan nilai di KPP Pratama Samarinda Ulu menggunakan metode K-NN diperoleh hasil yaitu dari 20 wajib pajak yang memiliki status patuh membayar PPN, terdapat 15 wajib pajak tepat diklasifikasikan memiliki status patuh membayar PPN dan 5 lainnya tidak tepat diklasifikasikan. Sedangkan dari 21 wajib pajak yang memiliki status tidak patuh membayar PPN, terdapat 18 wajib pajak tepat diklasifikasikan memiliki status tidak patuh membayar PPN dan 3

lainnya tidak tepat diklasifikasikan, sehingga berdasarkan Persamaan (2.11) maka diperoleh nilai APER sebagai berikut:

$$\text{APER} = \frac{5+3}{41} \times 100\%$$

$$= \frac{8}{41} \times 100\%$$

$$= 19,51\%$$

Jadi, nilai APER (laju *error*) pada hasil klasifikasi menggunakan 1-NN adalah sebesar 19,51%. Dengan cara yang sama peneliti juga melakukan prediksi menggunakan $K$ = 3,5,7 dan 9. Untuk menentukan nilai $K$ optimal, dapat dilakukan dengan melihat nilai APER dengan bantuan *software* R. Adapun nilai APER untuk seluruh nilai $K$ yang dicobakan dapat dilihat pada Tabel 4.9.

**Tabel 4.9** Nilai APER Pada Metode K-NN untuk Seluruh Nilai $K$

| **Nilai *K*** | **Laju *Error*** |
|---|---|
| 1 | 19,51% |
| 3 | 21,95% |
| 5 | 21,95% |
| 7 | 24,39% |
| 9 | 31,71% |

Sumber: Lampiran 9

Berdasarkan Tabel 4.9 dapat diketahui bahwa nilai APER yang paling kecil dihasilkan oleh $K$=1, sehingga dapat disimpulkan bahwa klasifikasi status pembayaran PPN menggunakan metode K-NN akan menghasilkan akurasi yang optimal apabila menggunakan $K$=1.

Selanjutnya dengan menggunakan seluruh nilai $K$, dilakukan prediksi kembali menggunakan data randomisasi yang lainnya. Adapun nilai APER pada hasil klasifikasi metode K-NN untuk semua percobaan randomisasi data disajikan pada Tabel 4.10 berikut:

**Tabel 4.10** Nilai APER Pada Metode K-NN untuk Setiap Percobaan Randomisasi Data

| **Nilai *K*** | **Randomisasi ke-** | | | | |
|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** |
| 1 | 36,59% | 31,71% | 26,83% | 19,51% | 29,27% |
| 3 | 31,71% | 31,71% | 26,83% | 21,95% | 21,95% |
| 5 | 31,71% | 29,27% | 29,26% | 21,95% | 21,95% |
| 7 | 31,71% | 24,39% | 24,39% | 24,39% | 24,39% |
| 9 | 31,71% | 31,71% | 36,59% | 31,71% | 31,71% |

Berdasarkan Tabel 4.10 dapat diketahui bahwa pada percobaan menggunakan 1-NN pada data randomisasi ke-4 mendapatkan nilai APER yang paling kecil, sehingga dapat disimpulkan dari semua percobaan yang dilakukan, percobaan 1-NN pada data randomisasi ke-4 dapat menghasilkan ketepatan klasifikasi metode K-NN yang paling akurat.

## 4.5 Perbandingan Tingkat Akurasi Klasifikasi

Pengukuran tingkat akurasi baik pada metode *naive* Bayes maupun K-NN dilakukan dengan menghitung peluang kesalahan klasifikasi. Pada proses klasifikasi diharapkan melakukan klasifikasi pada semua obyek dengan benar, sehingga semakin kecil kesalahan klasifikasi menunjukkan bahwa semakin baik hasil klasifikasi yang didapatkan. Pengukuran tingkat akurasi dilakukan berdasarkan hasil percobaan randomisasi data sebanyak 5 kali. Adapun hasil pengukuran tingkat akurasi terbaik pada kedua metode dengan menggunakan APER dapat dilihat pada Tabel 4.11.

**Tabel 4.11** Hasil Pengukuran Tingkat Akurasi Pada Metode Naive Bayes dan K-NN

| **Metode** | **APER** |
|---|---|
| *Naive* Bayes | 17,01% |
| *K-Nearest Neighbor* | 19,51% |

Pada Tabel 4.11 dapat diketahui bahwa pada metode *naive* Bayes menunjukkan kesalahan klasifikasi dalam memprediksi klasifikasi sebesar 17,01% dan pada metode K-NN menunjukkan kesalahan klasifikasi dalam memprediksi klasifikasi sebesar 19,51%. Hal ini menunjukkan bahwa metode *naive* Bayes bekerja lebih baik dibandingkan dengan metode K-NN dalam mengklasifikasikan status pembayaran PPN di KPP Pratama Samarinda Ulu dilihat dari nilai APER yang lebih rendah. Adapun hasil klasifikasi lengkap dapat dilihat pada Lampiran 8 dan Lampiran 11.

# BAB 5
# PENUTUP

## 5.1 Kesimpulan

Berdasarkan hasil analisis data dan pembahasan diperoleh kesimpulan sebagai berikut:

1. Pengklasifikasian status pembayaran PPN di KPP Pratama Samarinda Ulu tahun 2018 menggunakan metode *naive* Bayes diperoleh hasil yaitu dari 20 wajib pajak yang memiliki status patuh membayar PPN, terdapat 15 wajib pajak tepat diklasifikasikan memiliki status patuh membayar PPN dan 5 lainnya tidak tepat diklasifikasikan. Sedangkan dari 21 wajib pajak yang memiliki status tidak patuh membayar PPN, terdapat 19 wajib pajak tepat diklasifikasikan memiliki status tidak patuh membayar PPN dan 2 lainnya tidak tepat diklasifikasikan.
2. Pengklasifikasian status pembayaran PPN di KPP Pratama Samarinda Ulu tahun 2018 menggunakan metode K-NN diperoleh hasil yaitu 20 wajib pajak yang memiliki status patuh membayar PPN, terdapat 15 wajib pajak tepat diklasifikasikan memiliki status patuh membayar PPN dan 5 lainnya tidak tepat diklasifikasikan. Sedangkan dari 21 wajib pajak yang memiliki status tidak patuh membayar PPN, terdapat 18 wajib pajak tepat diklasifikasikan memiliki status tidak patuh membayar PPN dan 3 lainnya tidak tepat diklasifikasikan.
3. Hasil pengukuran tingkat akurasi klasifikasi menggunakan APER pada data status pembayaran PPN di KPP Pratama Samarinda Ulu tahun 2018 dengan metode *naive* Bayes diperoleh hasil kesalahan klasifikasi sebesar 17,07% dan metode K-NN sebesar 19,51%. Hal ini menunjukkan bahwa metode *naive* Bayes memberikan ketepatan prediksi klasifikasi yang lebih baik.

## 5.2 Saran

Pada penelitian ini menggunakan data kualitatif, sehingga dalam penelitian selanjutnya dapat menggunakan data kuantitatif pada variabel bebasnya dengan metode klasifikasi yang sama, khususnya untuk metode *naive* Bayes dapat

menggunakan pendekatan distribusi normal (gaussian) ataupun *kernel density estimation* (KDE). Selain itu, penelitian selanjutnya juga dapat menggunakan metode pengklasifikasian lainnya seperti *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM) atau *Probabilistic Neural Network* (PNN) serta dapat menerapkan metode *Adaptive Boosting* untuk memperbaiki tingkat akurasi pada hasil klasifikasi yang dibentuk.

# DAFTAR PUSTAKA


Antaristi, M. (2017). Aplikasi Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan Metode Naive Bayes di Bank BNI Syariah Surabaya. *Jurnal Teknik Elektro*, 9(2), 45-52.

Bustami. (2013). Penerapan Algoritma Naive Bayes untuk Mengklasifikasikan Data Nasabah Asuransi. *Jurnal Penelitian Teknik Informatika*, 3(2), 129-132.

Deza, M. dan Deza, L. (2009). *Encyclopedia of Distances*. New York: Springer.

Eko, R. (2014). Metode Klasifikasi Naive Bayes dan K-NN pada Analisis Data Status Kerja di Kabupaten Demak Tahun 2002. *Jurnal Gaussian*, 3(4), 831-838.

Han, J. dan Kamber, M. (2006). *Data Mining: Concepts and Techniques, Second Edition*. Burlington: Morgan Kaufmann Publishers.

Harrington, P. (2012). *Machine Learning in Action*. New York: Manning Publication.

John, G. dan Langley, P. (1995). *Estimating Continuous Distribution in Bayesian Classifiers*. San Mateo: Morgan Kaufmann Publishers.

Johnson, R. A. dan Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.

Judisseno, R. K. (2005). *Perpajakan (Edisi Revisi)*. Jakarta: Gramedia.

Kursini, L. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi.

Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons Inc.

Mardiasmo. (2005). *Akuntansi Sektor Publik*. Yogyakarta: Andi Offset.

Prahudaya, T. Y. (2017). Metode Klasifikasi Mutu Jambu Biji Menggunakan K-NN Berdasarkan Fitur Warna dan Tekstur. *Jurnal Ilmiah Sains dan Teknologi*, 6(2), 59-138.

Prasetyo, E. (2012). *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.

Prasetyo, E. (2014). *Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset.

Resmi, S. (2007). *Perpajakan Teori dan Kasus*. Jakarta: Salemba Empat.

Riduwan. (2013). *Skala Pengukuran Variabel-variabel Penelitian*. Bandung: Alfabeta.

Riduwan dan Akdon. (2010). *Rumus dan Data dalam Analisis Statistika Cetakan Kedua*. Bandung: Alfabeta.

Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.

Setia, P., Yuniarto, B., dan Mariyah, S. (2018). *Data Mining dengan R: Konsep Serta Implementasi*. Bogor: In Media.

Sugiyono. (2010). *Metode Penelitian Pendidikan Pendekatan Kualitatif, Kuantitatif, dan R&D*. Bandung: Alfabeta.

Towa, H. dan Tairas, J. N. B. (2002). *Pengantar Klasifikasi Persepuluhan Dewey*. Jakarta: BPK Gunung Mulya.

Turban, E., Aronson, J. E., dan Liang, T. P. (2005). *Decision Support Systems and Intelligent Systems*. Yogyakarta: Andi Offset.

Vulandari, R. T. (2017). *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta: Gava Media.

Waluyo. (2011). *Perpajakan Indonesia Edisi 10 Buku 1*. Jakarta: Salemba Empat.

Witten, I. H. dan Eibe, F. (2011). *Data Mining: Partical Machine Learning Tools and Techniques, Third Edition*. Burlington: Morgan Kaufmann Publishers.

# LAMPIRAN

**Lampiran 1.** Data Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 1 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 2 | Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 3 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 4 | Tidak Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 5 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 6 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 7 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 8 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 9 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 10 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 11 | Tidak Patuh | 100-500 Juta | Koperasi | Tepat Waktu |
| 12 | Patuh | 100-500 Juta | Koperasi | Tepat Waktu |
| 13 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 14 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 15 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 16 | Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 17 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 18 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 19 | Tidak Patuh | 100-500 Juta | Lainnya | Tepat Waktu |
| 20 | Tidak Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 21 | Patuh | < 100 Juta | Lainnya | Tidak Tepat Waktu |
| 22 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 23 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 24 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 25 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 26 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 27 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 28 | Tidak Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 29 | Tidak Patuh | 500 Juta - 1 Miliar | PT | Tidak Tepat Waktu |
| 30 | Patuh | 100-500 Juta | PT | Tepat Waktu |
| 31 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 32 | Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 33 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 34 | Tidak Patuh | < 100 Juta | PT | Tepat Waktu |
| 35 | Tidak Patuh | < 100 Juta | CV | Tepat Waktu |

**Lampiran 1.** Data Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 36 | Tidak Patuh | 1-10 Miliar | Koperasi | Tepat Waktu |
| 37 | Tidak Patuh | < 100 Juta | Koperasi | Tidak Tepat Waktu |
| 38 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 39 | Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 40 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 41 | Tidak Patuh | > 10 Miliar | PT | Tepat Waktu |
| 42 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 43 | Tidak Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 44 | Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 45 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 46 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 47 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 48 | Tidak Patuh | > 10 Miliar | CV | Tidak Tepat Waktu |
| 49 | Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 50 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 51 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 52 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 53 | Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |
| 54 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 55 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 56 | Tidak Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 57 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 58 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 59 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 60 | Patuh | 100-500 Juta | CV | Tepat Waktu |
| 61 | Patuh | < 100 Juta | Koperasi | Tidak Tepat Waktu |
| 62 | Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 63 | Tidak Patuh | 100-500 Juta | PT | Tepat Waktu |
| 64 | Patuh | > 10 Miliar | CV | Tepat Waktu |
| 65 | Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 66 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 67 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 68 | Tidak Patuh | < 100 Juta | Lainnya | Tidak Tepat Waktu |
| 69 | Tidak Patuh | 100-500 Juta | Lainnya | Tepat Waktu |
| 70 | Patuh | < 100 Juta | PT | Tepat Waktu |

**Lampiran 1.** Data Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 71 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 72 | Patuh | 1-10 Miliar | Lainnya | Tepat Waktu |
| 73 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 74 | Tidak Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 75 | Patuh | 100-500 Juta | PT | Tepat Waktu |
| 76 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 77 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 78 | Patuh | < 100 Juta | Koperasi | Tidak Tepat Waktu |
| 79 | Tidak Patuh | 100-500 Juta | Lainnya | Tidak Tepat Waktu |
| 80 | Patuh | 100-500 Juta | CV | Tepat Waktu |
| 81 | Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 82 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 83 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 84 | Tidak Patuh | < 100 Juta | Koperasi | Tepat Waktu |
| 85 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 86 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 87 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 88 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 89 | Tidak Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 90 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 91 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 92 | Tidak Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |
| 93 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 94 | Tidak Patuh | < 100 Juta | Lainnya | Tidak Tepat Waktu |
| 95 | Tidak Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 96 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 97 | Tidak Patuh | 500 Juta - 1 Miliar | Lainnya | Tepat Waktu |
| 98 | Tidak Patuh | 100-500 Juta | PT | Tepat Waktu |
| 99 | Patuh | > 10 Miliar | CV | Tepat Waktu |
| 100 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 101 | Patuh | 100-500 Juta | CV | Tidak Tepat Waktu |
| 102 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 103 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 104 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 105 | Patuh | < 100 Juta | CV | Tepat Waktu |

**Lampiran 1.** Data Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 106 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 107 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 108 | Tidak Patuh | 1-10 Miliar | Koperasi | Tepat Waktu |
| 109 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 110 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 111 | Tidak Patuh | > 10 Miliar | Lainnya | Tepat Waktu |
| 112 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 113 | Patuh | 1-10 Miliar | Lainnya | Tepat Waktu |
| 114 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 115 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 116 | Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 117 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 118 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 119 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 120 | Tidak Patuh | > 10 Miliar | PT | Tepat Waktu |
| 121 | Patuh | 1-10 Miliar | Lainnya | Tepat Waktu |
| 122 | Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |
| 123 | Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 124 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 125 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 126 | Patuh | 100-500 Juta | CV | Tidak Tepat Waktu |
| 127 | Tidak Patuh | < 100 Juta | Lainnya | Tidak Tepat Waktu |
| 128 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 129 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 130 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 131 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 132 | Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 133 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 134 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 135 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 136 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 137 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 138 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 139 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 140 | Tidak Patuh | 100-500 Juta | PT | Tidak Tepat Waktu |

**Lampiran 1.** Data Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 141 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 142 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 143 | Tidak Patuh | < 100 Juta | Koperasi | Tidak Tepat Waktu |
| 144 | Tidak Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |
| 145 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 146 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 147 | Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 148 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 149 | Patuh | 500 Juta - 1 Miliar | Lainnya | Tepat Waktu |
| 150 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 151 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 152 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 153 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 154 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 155 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 156 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 157 | Tidak Patuh | 100-500 Juta | PT | Tepat Waktu |
| 158 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 159 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 160 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 161 | Tidak Patuh | > 10 Miliar | CV | Tepat Waktu |
| 162 | Tidak Patuh | 100-500 Juta | CV | Tidak Tepat Waktu |
| 163 | Patuh | > 10 Miliar | CV | Tepat Waktu |
| 164 | Tidak Patuh | 100-500 Juta | CV | Tidak Tepat Waktu |
| 165 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 166 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 167 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 168 | Patuh | 100-500 Juta | CV | Tepat Waktu |
| 169 | Tidak Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 170 | Tidak Patuh | < 100 Juta | CV | Tepat Waktu |
| 171 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 172 | Tidak Patuh | > 10 Miliar | CV | Tepat Waktu |
| 173 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 174 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 175 | Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |

**Lampiran 1.** Data Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 176 | Patuh | 500 Juta - 1 Miliar | CV | Tidak Tepat Waktu |
| 177 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 178 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 179 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 180 | Tidak Patuh | 1-10 Miliar | Koperasi | Tidak Tepat Waktu |
| 181 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 182 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 183 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 184 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 185 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 186 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 187 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 188 | Tidak Patuh | > 10 Miliar | PT | Tepat Waktu |
| 189 | Tidak Patuh | 100-500 Juta | Koperasi | Tepat Waktu |
| 190 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 191 | Tidak Patuh | 1-10 Miliar | Koperasi | Tidak Tepat Waktu |
| 192 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 193 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 194 | Patuh | > 10 Miliar | CV | Tepat Waktu |
| 195 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 196 | Patuh | 1-10 Miliar | Koperasi | Tepat Waktu |
| 197 | Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 198 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 199 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 200 | Tidak Patuh | > 10 Miliar | Koperasi | Tidak Tepat Waktu |
| 201 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 202 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 203 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 204 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 205 | Tidak Patuh | > 10 Miliar | Koperasi | Tidak Tepat Waktu |

**Lampiran 2.** *Syntax Pie Chart* dengan Program R

```
> windowsFonts(A=windowsFont("Times New Roman"))
> Data= read.table("E://DataWP.txt",header=TRUE)

#Pie Chart Variabel Status Pembayaran PPN
> slices<-c(103,102)
> lbls<-c("Patuh","Tidak Patuh")
> pct<-round(slices/sum(slices)*100)
> lbls<-paste(lbls,pct)
> lbls<-paste(lbls,"%",sep="")
> pie(slices,labels=lbls,family="A",col=rainbow(length(lbls)),main="Status
  Pembayaran PPN")

#Pie Chart Variabel Pendapatan
> slices<-c(80,30,21,49,25)
> lbls<-c("<100Jt","100-500Jt","500Jt-1M","1-10M",">10M")
> pct<-round(slices/sum(slices)*100)
> lbls<-paste(lbls,pct)
> lbls<-paste(lbls,"%",sep="")
> pie(slices,labels=lbls,family="A",col=rainbow(length(lbls)),main="Penda
  patan")

#Pie Chart Variabel Bentuk Badan
> slices<-c(74,87,9,15,20)
> lbls<-c("PT","CV""","BUMN/BUMD","Koperasi","Lainnya")
> pct<-round(slices/sum(slices)*100)
> lbls<-paste(lbls,pct)
> lbls<-paste(lbls,"%",sep="")
> pie(slices,labels=lbls,family="A",col=rainbow(length(lbls)),main="Bentuk
  Badan")

#Pie Chart Variabel Status Pelaporan Pajak
> slices<-c(144,61)
> lbls<-c("Tepat Waktu","Tidak Tepat Waktu")
> pct<-round(slices/sum(slices)*100)
> lbls<-paste(lbls,pct)
> lbls<-paste(lbls,"%",sep="")
> pie(slices,labels=lbls,family="A",col=rainbow(length(lbls)),main="Status
  Pelaporan Pajak")
```

**Lampiran 3.** *Syntax* dan *Output* Randomisasi Data dengan Program R

```
#Percobaan Pertama
> Data= read.table("E://DataWP.txt",header=TRUE)
> Random=DataWP[(sample(1:205)]
> Random
 [1]  80 192  42 170 115  83 173 171  59 197 196  82  76 104 149 108 166  54
 [19] 13   3 129 148 167  22 122 131  56 145  47  78 169  81  20  67 174  92
 [37] 30  34  21  75 199  40  58  12  74 195 184  18 180 152  70 135 127  45
 [55] 51 141  57 185 163 116 150 154 103 147 119  41 146 178 136 100 189 177
 [73] 95 138 156  63 137 164 121 175 118 142   5  86 101 105  10 132 153  35
 [91] 36  27 125  60  98  97 120 193  96 187  68 155  61  66 113  89 124  50
[109] 168 16 157  32 140  62 162 106 158  49  99  37  26  90  71 194   4 200
[127] 201  69  65 203 190   7 139   9  79  19  52 191 151 160 165 202   2 134
[145]  29 179  25 186  55  11  64 183 111 38 110  14  72 109 143  8  130 126
[163] 128 172  84  87  31  85  44  93  77   6 204  94  46 123  23 102  48  17
[181] 144 181 198 114 117 159 205 53 112  43   1  88 176  39  91 133 182  15
[199]  73 188  24  28 107  33 161

#Percobaan Kedua
> Data= read.table("E://DataWP.txt",header=TRUE)
> Random=DataWP[(sample(1:205)]
> Random
 [1] 159 141  27 182  42  77 100 161 158 191  99  79  92  15 170 118 181 127
 [19] 160 121  60  24  10   6 174  90  84 103 138 165  85 192  95 112 104  28
 [37] 194 176  76  89  58   1  26 106 142   4 186  52 140  66 203  19  86  43
 [55]  87 180  54 144 148 150 185  38 143  61  72 108 145   9 187 195  36  20
 [73]  17 147  13 129 183 109   3 184 201  57 116   7  97 151 119 200  22  74
 [91] 122  98  18  12 135 115 117 134 114  56 175 156 162  71  14   5 169  75
[109]  46  68  62  91 139  83 152   2 204  70  94  88 120 171 136 124  39 132
[127] 102 137 202  63 130  48 198  50  25 133 154  81  37  33  49 110  30 157
[145]  35  16  65  44 177 193  93 128 178 173 149 78 146 164 126  34  69  47
[163] 101 105  51  80 107 196 190 197 199 73 113 153 155  59 123 163  31 131
[181]  23  55  53  40   8 179 172  11 168  45 111 188  96  64  32 205 166  21
[199]  41 167  67 189 125  82  29

#Percobaan Ketiga
> Data= read.table("E://DataWP.txt",header=TRUE)
> Random=DataWP[(sample(1:205)]
> Random
```

**Lampiran 3.** *Syntax* dan *Output* Randomisasi Data dengan Program R (Lanjutan)

```
 [1]   16  157   7   111  58  154  50  196 162 174 159 141 167 78  202   1   29  117
 [19] 169 156 194 120 149  74  197  46  178 104   4  133 103 173 132 150 187  39
 [37] 105  53  172  60  153 176 59  146   34  75   61 136  26  122 144  97  177  44
 [55]  99  41  175 190  138 198  28 121  179 183  84   73  37   3    163  27  113  56
 [73]  36 195 204  89   181 102 124 96   90  135 180   2   10  205 127  71   52   54
 [91]  38  21  17  107   88  142  32   63  11  100 189  51  15   42  106  143 110 155
[109] 166 140 165 70 164  94  48  76  168  19   86  137 125 115 191   9   108 139
[127]   5  35 134  40  109 145 147  31  23  93  200  67  192  30  114  85   126  98
[145] 171  24  20 158 185 188  14 123  92 193  18  80    91  77   148  66    62  83
[163] 161 160  87  57 152  45  72 112 119  47  203  25   43  68   129  131  201  95
[181]  65  69   8  199  116  82 182  64 118  13  79 151  33  55    49   128   81  184
[199]  22 186 101  6   130  12  170

#Percobaan Keempat
> Data= read.table("E://DataWP.txt",header=TRUE)
> Random=DataWP[(sample(1:205)]
> Random
 [1]  105  69  107 112  111  178  47  7   30   24 130   22   167  97  40  34  155   91
 [19] 28  146  154  95 168   61   5  144 78  180  79  170  122  87 174 123 157   2
 [37] 19   92  131  16    31   57  86 181  37  84 141  137  182  94  25  126 151 119
 [55]  4  198   46   12  185  75  53 108  26  158 203  85   139 110 100 20  113  162
 [73] 29   9   183 169  202  147 33 116 133 106 190 143 166  65   89  50   70  72
 [91] 98  27  120  52   176  66  60 140  82  128 134 164   36  188  55 18  186  136
[109] 163 76  42  21   121  48  11 103  74  191  58  104  127 142  10 193 117  200
[127] 161 83 160 173   6   187 32 165  96 109 199   43    8    38  175 148 184 124
[145]  145 201 179 64 101 102  68 56  71  54  197   63   39   88  171  172  35  138
[163]  90   3   177  51  23   77 125 118 114 115 14  80  159  49   13  192   1   194
[181]  152  93  73  44  59  132  41 196 156  99 195 153 150  81  204  62   67   45
[199]  15  205 149 189 129 135 17

#Percobaan Kelima
> Data= read.table("E://DataWP.txt",header=TRUE)
> Random=DataWP[(sample(1:205)]
> Random
```

**Lampiran 3.** *Syntax* dan *Output* Randomisasi Data dengan Program R (Lanjutan)

```
[1]   71  97   47  152   2  190  102  110  27  82  179 154 118  67 101  70  98   4
[19] 120 121 177 108  12  200  195  49   62  189 133  52  17  55  193  16  44  181
[37] 125 94  201  21  174  10   34  159  46  134  205 188  89  59  18 146 149 147
[55]  68 176  58  84  115 123   63   96  103  79  170   91 104  39  40 167 109  86
[73] 165  41  75  88  124  42   194  169 127 135  33   78 129 144  36 168  173  8
[91]  64  32  76  171  203 183  72    1   99  116 132 143  61  87  130  19  202  92
[109] 158 131 137 117 60  53  81 160  151  145 119  50    6   204  69 187 164 153
[127]   5  37   9  172   14  35   90 100 148 192  198  112  66  140 142  13 128 184
[145] 106  28 180 156  20  30 166 122  83   22    26    45 126   73 178  77  111  23
[163] 162 107  56 105 197  57  43 157 136  25  155   65 150  114  38 199 191 139
[181]  51   3   185 74  161  95  15 113   85  54   31    7  141  163  138  11  80 182
[199] 186  29  48  93  175  24 196
```

**Lampiran 4.** Data Randomisasi Percobaan Keempat Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 105 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 69 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 107 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 112 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 111 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 178 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 47 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 7 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 30 | Patuh | 100-500 Juta | PT | Tepat Waktu |
| 24 | Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 130 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 22 | Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 167 | Tidak Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 97 | Tidak Patuh | < 100 Juta | CV | Tepat Waktu |
| 40 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 34 | Patuh | 100-500 Juta | Koperasi | Tepat Waktu |
| 155 | Tidak Patuh | 100-500 Juta | Koperasi | Tepat Waktu |
| 91 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 28 | Patuh | < 100 Juta | Lainnya | Tidak Tepat Waktu |
| 146 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 154 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 95 | Tidak Patuh | < 100 Juta | PT | Tepat Waktu |
| 168 | Tidak Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 61 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 5 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 144 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 78 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 180 | Tidak Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 79 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 170 | Tidak Patuh | 500 Juta - 1 Miliar | PT | Tidak Tepat Waktu |
| 122 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 87 | Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 174 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 123 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 157 | Tidak Patuh | 100-500 Juta | Lainnya | Tepat Waktu |

**Lampiran 4.** Data Randomisasi Percobaan Keempat Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 2 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 19 | Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 92 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 131 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 16 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 31 | Patuh | 100-500 Juta | CV | Tepat Waktu |
| 57 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 86 | Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 181 | Tidak Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 37 | Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |
| 84 | Patuh | > 10 Miliar | CV | Tepat Waktu |
| 141 | Tidak Patuh | 100-500 Juta | PT | Tepat Waktu |
| 137 | Tidak Patuh | < 100 Juta | Koperasi | Tidak Tepat Waktu |
| 182 | Tidak Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 94 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 25 | Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 126 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 151 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 119 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 4 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 198 | Tidak Patuh | > 10 Miliar | PT | Tepat Waktu |
| 46 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 12 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 185 | Tidak Patuh | 1-10 Miliar | Koperasi | Tepat Waktu |
| 75 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 53 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 108 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 26 | Patuh | < 100 Juta | Koperasi | Tidak Tepat Waktu |
| 158 | Tidak Patuh | 100-500 Juta | Lainnya | Tepat Waktu |
| 203 | Tidak Patuh | > 10 Miliar | CV | Tidak Tepat Waktu |
| 85 | Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 139 | Tidak Patuh | < 100 Juta | Lainnya | Tidak Tepat Waktu |
| 110 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 100 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 20 | Patuh | < 100 Juta | Lainnya | Tepat Waktu |

**Lampiran 4.** Data Randomisasi Percobaan Keempat Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 113 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 162 | Tidak Patuh | 100-500 Juta | Lainnya | Tidak Tepat Waktu |
| 29 | Patuh | 100-500 Juta | PT | Tepat Waktu |
| 9 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 183 | Tidak Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 169 | Tidak Patuh | 500 Juta - 1 Miliar | Lainnya | Tepat Waktu |
| 202 | Tidak Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 147 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 33 | Patuh | 100-500 Juta | CV | Tepat Waktu |
| 116 | Tidak Patuh | < 100 Juta | Koperasi | Tepat Waktu |
| 133 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 106 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 190 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 143 | Tidak Patuh | 100-500 Juta | PT | Tepat Waktu |
| 166 | Tidak Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 65 | Patuh | 1-10 Miliar | Lainnya | Tepat Waktu |
| 89 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 50 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 70 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 72 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 98 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 27 | Patuh | < 100 Juta | Koperasi | Tidak Tepat Waktu |
| 120 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 52 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 176 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 66 | Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 60 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 140 | Tidak Patuh | < 100 Juta | Lainnya | Tidak Tepat Waktu |
| 82 | Patuh | > 10 Miliar | CV | Tepat Waktu |
| 128 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 134 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 164 | Tidak Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |
| 36 | Patuh | 100-500 Juta | CV | Tidak Tepat Waktu |
| 188 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 55 | Patuh | 1-10 Miliar | PT | Tepat Waktu |

**Lampiran 4.** Data Randomisasi Percobaan Keempat Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 18 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 186 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 136 | Tidak Patuh | < 100 Juta | Koperasi | Tidak Tepat Waktu |
| 163 | Tidak Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |
| 76 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 42 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 21 | Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 121 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 48 | Patuh | 500 Juta - 1 Miliar | Lainnya | Tepat Waktu |
| 11 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 103 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 74 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 191 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 58 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 104 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 127 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 142 | Tidak Patuh | 100-500 Juta | PT | Tepat Waktu |
| 10 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 193 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 117 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 200 | Tidak Patuh | > 10 Miliar | CV | Tepat Waktu |
| 161 | Tidak Patuh | 100-500 Juta | CV | Tidak Tepat Waktu |
| 83 | Patuh | > 10 Miliar | CV | Tepat Waktu |
| 160 | Tidak Patuh | 100-500 Juta | CV | Tidak Tepat Waktu |
| 173 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 6 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 187 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 32 | Patuh | 100-500 Juta | CV | Tepat Waktu |
| 165 | Tidak Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 96 | Tidak Patuh | < 100 Juta | CV | Tepat Waktu |
| 109 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 199 | Tidak Patuh | > 10 Miliar | CV | Tepat Waktu |
| 43 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 8 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 38 | Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |

**Lampiran 4.** Data Randomisasi Percobaan Keempat Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 175 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 148 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 184 | Tidak Patuh | 1-10 Miliar | Koperasi | Tepat Waktu |
| 124 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 145 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 201 | Tidak Patuh | > 10 Miliar | Lainnya | Tepat Waktu |
| 179 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 64 | Patuh | 1-10 Miliar | Lainnya | Tepat Waktu |
| 101 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 102 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 68 | Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 56 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 71 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 54 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 197 | Tidak Patuh | > 10 Miliar | PT | Tepat Waktu |
| 63 | Patuh | 1-10 Miliar | Lainnya | Tepat Waktu |
| 39 | Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |
| 88 | Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 171 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 172 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 35 | Patuh | 100-500 Juta | CV | Tidak Tepat Waktu |
| 138 | Tidak Patuh | < 100 Juta | Lainnya | Tidak Tepat Waktu |
| 90 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 3 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 177 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 51 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 23 | Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 77 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 125 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 118 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 114 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 115 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 14 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 80 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 159 | Tidak Patuh | 100-500 Juta | PT | Tidak Tepat Waktu |

**Lampiran 4.** Data Randomisasi Percobaan Keempat Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 49 | Patuh | 500 Juta - 1 Miliar | CV | Tidak Tepat Waktu |
| 13 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 192 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 1 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 194 | Tidak Patuh | 1-10 Miliar | Koperasi | Tidak Tepat Waktu |
| 152 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 93 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 73 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 44 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 59 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 132 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 41 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 196 | Tidak Patuh | > 10 Miliar | PT | Tepat Waktu |
| 156 | Tidak Patuh | 100-500 Juta | Koperasi | Tepat Waktu |
| 99 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 195 | Tidak Patuh | 1-10 Miliar | Koperasi | Tidak Tepat Waktu |
| 153 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 150 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 81 | Patuh | > 10 Miliar | CV | Tepat Waktu |
| 204 | Tidak Patuh | > 10 Miliar | Koperasi | Tidak Tepat Waktu |
| 62 | Patuh | 1-10 Miliar | Koperasi | Tepat Waktu |
| 67 | Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 45 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 15 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 205 | Tidak Patuh | > 10 Miliar | Koperasi | Tidak Tepat Waktu |
| 149 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 189 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 129 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 135 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 17 | Patuh | < 100 Juta | CV | Tepat Waktu |

**Lampiran 5.** Data *Training* Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 105 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 69 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 107 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 112 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 111 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 178 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 47 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 7 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 30 | Patuh | 100-500 Juta | PT | Tepat Waktu |
| 24 | Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 130 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 22 | Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 167 | Tidak Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 97 | Tidak Patuh | < 100 Juta | CV | Tepat Waktu |
| 40 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 34 | Patuh | 100-500 Juta | Koperasi | Tepat Waktu |
| 155 | Tidak Patuh | 100-500 Juta | Koperasi | Tepat Waktu |
| 91 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 28 | Patuh | < 100 Juta | Lainnya | Tidak Tepat Waktu |
| 146 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 154 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 95 | Tidak Patuh | < 100 Juta | PT | Tepat Waktu |
| 168 | Tidak Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 61 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 5 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 144 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 78 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 180 | Tidak Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 79 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 170 | Tidak Patuh | 500 Juta - 1 Miliar | PT | Tidak Tepat Waktu |
| 122 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 87 | Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 174 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 123 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 157 | Tidak Patuh | 100-500 Juta | Lainnya | Tepat Waktu |

**Lampiran 5.** Data *Training* Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 2 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 19 | Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 92 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 131 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 16 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 31 | Patuh | 100-500 Juta | CV | Tepat Waktu |
| 57 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 86 | Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 181 | Tidak Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 37 | Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |
| 84 | Patuh | > 10 Miliar | CV | Tepat Waktu |
| 141 | Tidak Patuh | 100-500 Juta | PT | Tepat Waktu |
| 137 | Tidak Patuh | < 100 Juta | Koperasi | Tidak Tepat Waktu |
| 182 | Tidak Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 94 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 25 | Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 126 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 151 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 119 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 4 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 198 | Tidak Patuh | > 10 Miliar | PT | Tepat Waktu |
| 46 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 12 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 185 | Tidak Patuh | 1-10 Miliar | Koperasi | Tepat Waktu |
| 75 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 53 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 108 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 26 | Patuh | < 100 Juta | Koperasi | Tidak Tepat Waktu |
| 158 | Tidak Patuh | 100-500 Juta | Lainnya | Tepat Waktu |
| 203 | Tidak Patuh | > 10 Miliar | CV | Tidak Tepat Waktu |
| 85 | Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 139 | Tidak Patuh | < 100 Juta | Lainnya | Tidak Tepat Waktu |
| 110 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 100 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 20 | Patuh | < 100 Juta | Lainnya | Tepat Waktu |

**Lampiran 5.** Data *Training* Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 113 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 162 | Tidak Patuh | 100-500 Juta | Lainnya | Tidak Tepat Waktu |
| 29 | Patuh | 100-500 Juta | PT | Tepat Waktu |
| 9 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 183 | Tidak Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 169 | Tidak Patuh | 500 Juta - 1 Miliar | Lainnya | Tepat Waktu |
| 202 | Tidak Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 147 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 33 | Patuh | 100-500 Juta | CV | Tepat Waktu |
| 116 | Tidak Patuh | < 100 Juta | Koperasi | Tepat Waktu |
| 133 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 106 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 190 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 143 | Tidak Patuh | 100-500 Juta | PT | Tepat Waktu |
| 166 | Tidak Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 65 | Patuh | 1-10 Miliar | Lainnya | Tepat Waktu |
| 89 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 50 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 70 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 72 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 98 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 27 | Patuh | < 100 Juta | Koperasi | Tidak Tepat Waktu |
| 120 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 52 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 176 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 66 | Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 60 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 140 | Tidak Patuh | < 100 Juta | Lainnya | Tidak Tepat Waktu |
| 82 | Patuh | > 10 Miliar | CV | Tepat Waktu |
| 128 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 134 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 164 | Tidak Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |
| 36 | Patuh | 100-500 Juta | CV | Tidak Tepat Waktu |
| 188 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 55 | Patuh | 1-10 Miliar | PT | Tepat Waktu |

**Lampiran 5.** Data *Training* Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 18 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 186 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 136 | Tidak Patuh | < 100 Juta | Koperasi | Tidak Tepat Waktu |
| 163 | Tidak Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |
| 76 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 42 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 21 | Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 121 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 48 | Patuh | 500 Juta - 1 Miliar | Lainnya | Tepat Waktu |
| 11 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 103 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 74 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 191 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 58 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 104 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 127 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 142 | Tidak Patuh | 100-500 Juta | PT | Tepat Waktu |
| 10 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 193 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 117 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 200 | Tidak Patuh | > 10 Miliar | CV | Tepat Waktu |
| 161 | Tidak Patuh | 100-500 Juta | CV | Tidak Tepat Waktu |
| 83 | Patuh | > 10 Miliar | CV | Tepat Waktu |
| 160 | Tidak Patuh | 100-500 Juta | CV | Tidak Tepat Waktu |
| 173 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 6 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 187 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 32 | Patuh | 100-500 Juta | CV | Tepat Waktu |
| 165 | Tidak Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 96 | Tidak Patuh | < 100 Juta | CV | Tepat Waktu |
| 109 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 199 | Tidak Patuh | > 10 Miliar | CV | Tepat Waktu |
| 43 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 8 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 38 | Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |

**Lampiran 5.** Data *Training* Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 175 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 148 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 184 | Tidak Patuh | 1-10 Miliar | Koperasi | Tepat Waktu |
| 124 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 145 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 201 | Tidak Patuh | > 10 Miliar | Lainnya | Tepat Waktu |
| 179 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 64 | Patuh | 1-10 Miliar | Lainnya | Tepat Waktu |
| 101 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 102 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 68 | Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 56 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 71 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 54 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 197 | Tidak Patuh | > 10 Miliar | PT | Tepat Waktu |
| 63 | Patuh | 1-10 Miliar | Lainnya | Tepat Waktu |
| 39 | Patuh | 500 Juta - 1 Miliar | PT | Tepat Waktu |
| 88 | Patuh | > 10 Miliar | PT | Tidak Tepat Waktu |
| 171 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 172 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 35 | Patuh | 100-500 Juta | CV | Tidak Tepat Waktu |
| 138 | Tidak Patuh | < 100 Juta | Lainnya | Tidak Tepat Waktu |
| 90 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 3 | Patuh | < 100 Juta | PT | Tepat Waktu |

**Lampiran 6.** Data *Testing* Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 177 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 51 | Patuh | 1-10 Miliar | PT | Tepat Waktu |
| 23 | Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 77 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 125 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu |
| 118 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu |
| 114 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 115 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu |
| 14 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 80 | Patuh | > 10 Miliar | PT | Tepat Waktu |
| 159 | Tidak Patuh | 100-500 Juta | PT | Tidak Tepat Waktu |
| 49 | Patuh | 500 Juta - 1 Miliar | CV | Tidak Tepat Waktu |
| 13 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 192 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 1 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 194 | Tidak Patuh | 1-10 Miliar | Koperasi | Tidak Tepat Waktu |
| 152 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 93 | Patuh | < 100 Juta | PT | Tepat Waktu |
| 73 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu |
| 44 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 59 | Patuh | 1-10 Miliar | CV | Tepat Waktu |
| 132 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 41 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 196 | Tidak Patuh | > 10 Miliar | PT | Tepat Waktu |
| 156 | Tidak Patuh | 100-500 Juta | Koperasi | Tepat Waktu |
| 99 | Patuh | < 100 Juta | CV | Tepat Waktu |
| 195 | Tidak Patuh | 1-10 Miliar | Koperasi | Tidak Tepat Waktu |
| 153 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 150 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 81 | Patuh | > 10 Miliar | CV | Tepat Waktu |
| 204 | Tidak Patuh | > 10 Miliar | Koperasi | Tidak Tepat Waktu |
| 62 | Patuh | 1-10 Miliar | Koperasi | Tepat Waktu |
| 67 | Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 45 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu |
| 15 | Patuh | < 100 Juta | CV | Tepat Waktu |

**Lampiran 6.** Data *Testing* Wajib Pajak Badan di KPP Pratama Samarinda Ulu Tahun 2018 (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak |
|---|---|---|---|---|
| 205 | Tidak Patuh | > 10 Miliar | Koperasi | Tidak Tepat Waktu |
| 149 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu |
| 189 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu |
| 129 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 135 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu |
| 17 | Patuh | < 100 Juta | CV | Tepat Waktu |

**Lampiran 7.** *Syntax Naive* Bayes dengan Program R

```
#Menggunakan Data Random Pertama
> library(e1071)
> Data= read.table("E://DataWP1.txt",header=TRUE)
> Data$Y<-factor(Data$Y)
> Data$X1<-factor(Data$X1)
> Data$X2<-factor(Data$X2)
> Data$X3<-factor(Data$X3)
> DataTraining<-Data[1:164,]
> DataTesting<-Data[165:205,]
> testing=Data[165:205,1]
> NB<-naiveBayes(Y~X1+X2+X3,data=DataTraining)
> NB
Naive Bayes Classifier for Discrete Predictors
Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)
A-priori probabilities:
Y
   Patuh      TidakPatuh
0.5243902    0.4756098
Conditional probabilities:
                                        X1
    Y              <100 Jt     100-500 Jt   500Jt-1M       1-10M         >10M
  Patuh          0.39534884 0.08139535 0.12790698 0.25581395 0.13953488
  Tidak Patuh   0.38461538 0.21794872 0.06410256 0.21794872 0.11538462
                                        X2
    Y                PT            CV    BUMN/BUMD      KOP       Lainnya
  Patuh          0.41860465 0.48837209 0.00000000 0.03488372 0.05813953
  Tidak Patuh   0.30769231 0.38461538 0.07692308 0.08974359 0.14102564X3
    Y          Tepat Waktu  Tidak Tepat Waktu
  Patuh          0.7906977 0.2093023
  Tidak Patuh   0.6025641 0.3974359

> Prediksi<-predict(NB,DataTesting,type=c("class"))
> Prediksi
 [1] Tidak Patuh  Tidak Patuh   Patuh         Tidak Patuh  Tidak Patuh  Tidak Patuh
[7] Patuh          Patuh          Patuh        Tidak Patuh   Patuh         Patuh
[13] Tidak Patuh Tidak Patuh   Patuh         Patuh          Patuh         Tidak Patuh
[19] Tidak Patuh Tidak Patuh   Patuh         Tidak Patuh   Tidak Patuh Tidak Patuh
[25] Patuh         Tidak Patuh   Patuh         Tidak Patuh   Patuh         Patuh
[31] Patuh         Tidak Patuh   Patuh         Patuh         Tidak Patuh  Tidak Patuh
[37] Patuh         Tidak Patuh   Tidak Patuh Tidak Patuh  Tidak Patuh
Levels: Patuh  Tidak Patuh
```

**Lampiran 7.** *Syntax Naive* Bayes dengan Program R (Lanjutan)

```
> MatriksKonfusi=table(testing,Prediksi)
> MatriksKonfusi
                    Prediksi
testing          Patuh  Tidak  Patuh
    Patuh          11        6
    Tidak  Patuh    7        17
> Akurasi=100*sum(testing==Prediksi)/NROW(testing)
> Akurasi
[1] 68.29268
> Laju_error=100*(MatriksKonfusi[2]+MatriksKonfusi[3])/NROW(testing)
> Laju_error
[1] 31.70732


#Menggunakan  Data Random Kedua
> library(e1071)
> Data= read.table("E://DataWP2.txt",header=TRUE)
> Data$Y<-factor(Data$Y)
> Data$X1<-factor(Data$X1)
> Data$X2<-factor(Data$X2)
> Data$X3<-factor(Data$X3)
> DataTraining<-Data[1:164,]
> DataTesting<-Data[165:205,]
> testing=Data[165:205,1]
> NB<-naiveBayes(Y~X1+X2+X3,data=DataTraining)
> NB
Naive Bayes Classifier  for Discrete  Predictors
Call:
naiveBayes.default(x  = X, y = Y, laplace  = laplace)
A-priori probabilities:
Y
   Patuh       TidakPatuh
0.5060976    0.4939024
Conditional  probabilities:
                                        X1
    Y              <100 Jt     100-500 Jt   500Jt-1M       1-10M         >10M
  Patuh          0.42168675 0.06024096 0.12048193 0.24096386 0.15662651
  Tidak  Patuh   0.34567901 0.20987654 0.09876543 0.23456790 0.11111111
                                        X2
    Y                PT            CV    BUMN/BUMD      KOP       Lainnya
  Patuh          0.44578313 0.43373494 0.00000000 0.04819277 0.07228916
  Tidak  Patuh   0.34567901 0.34567901 0.06172840 0.12345679 0.12345679
```

**Lampiran 7.** *Syntax Naive* Bayes dengan Program R (Lanjutan)

```
                                        X3
   Y          Tepat Waktu  Tidak Tepat Waktu
 Patuh          0.8072289 0.1927711
 Tidak Patuh   0.6049383 0.3950617

> Prediksi<-predict(NB,DataTesting,type=c("class"))
> Prediksi
[1] Tidak Patuh  Tidak Patuh  Tidak Patuh  Tidak Patuh  Tidak Patuh  Tidak Patuh
[7] Patuh        Patuh        Tidak Patuh  Patuh        Patuh        Tidak Patuh
[13] Tidak Patuh Patuh        Tidak Patuh  Tidak Patuh  Tidak Patuh  Tidak Patuh
[19] Tidak Patuh Tidak Patuh  Tidak Patuh  Patuh        Tidak Patuh  Tidak Patuh
[25] Tidak Patuh  Tidak Patuh  Patuh        Patuh        Patuh        Patuh
[31] Tidak Patuh  Patuh        Patuh        Patuh        Tidak Patuh  Tidak Patuh
[37] Tidak Patuh  Tidak Patuh  Patuh        Tidak Patuh  Tidak Patuh
Levels: Patuh  Tidak Patuh

> MatriksKonfusi=table(testing,Prediksi)
> MatriksKonfusi
                    Prediksi
testing          Patuh  Tidak Patuh
    Patuh          11        9
    Tidak Patuh     3        18
> Akurasi=100*sum(testing==Prediksi)/NROW(testing)
> Akurasi
[1] 70.73171
> Laju_error=100*(MatriksKonfusi[2]+MatriksKonfusi[3])/NROW(testing)
> Laju_error
[1] 29.26829


#Menggunakan Data Random Ketiga
> library(e1071)
> Data= read.table("E://DataWP3.txt",header=TRUE)
> Data$Y<-factor(Data$Y)
> Data$X1<-factor(Data$X1)
> Data$X2<-factor(Data$X2)
> Data$X3<-factor(Data$X3)
> DataTraining<-Data[1:164,]
> DataTesting<-Data[165:205,]
> testing=Data[165:205,1]
> NB<-naiveBayes(Y~X1+X2+X3,data=DataTraining)
> NB
```

**Lampiran 7.** *Syntax Naive* Bayes dengan Program R (Lanjutan)

```
Naive Bayes Classifier for Discrete Predictors
Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)
A-priori probabilities:
Y
   Patuh      TidakPatuh
0.4817073    0.5182927
Conditional probabilities:
                                        X1
    Y              <100 Jt     100-500 Jt   500Jt-1M       1-10M         >10M
  Patuh          0.41772152 0.07594937 0.15189873 0.24050633 0.11392405
  Tidak Patuh   0.30588235 0.25882353 0.08235294 0.23529412 0.11764706
                                        X2
    Y                PT            CV    BUMN/BUMD      KOP       Lainnya
  Patuh          0.40506329 0.45569620 0.00000000 0.05063291 0.08860759
  Tidak Patuh   0.31764706 0.42352941 0.05882353 0.08235294 0.11764706
                                        X3
    Y          Tepat Waktu  Tidak Tepat Waktu
  Patuh          0.7594937 0.2405063
  Tidak Patuh   0.6352941 0.3647059

> Prediksi<-predict(NB,DataTesting,type=c("class"))
> Prediksi
 [1] Patuh          Tidak Patuh   Patuh          Patuh          Patuh          Patuh
[7] Tidak Patuh   Tidak Patuh   Patuh          Tidak Patuh   Patuh          Patuh
[13] Patuh          Patuh          Patuh          Patuh          Tidak Patuh   Patuh
[19] Patuh          Patuh          Patuh          Patuh          Tidak Patuh   Tidak Patuh
[25] Patuh          Tidak Patuh Patuh          Patuh          Tidak Patuh   Tidak Patuh
[31] Patuh          Tidak Patuh Patuh          Patuh          Tidak Patuh   Tidak Patuh
[37] Tidak Patuh   Patuh          Tidak Patuh   Tidak Patuh   Patuh
Levels: Patuh   Tidak Patuh

> MatriksKonfusi=table(testing,Prediksi)
> MatriksKonfusi
                        Prediksi
testing              Patuh  Tidak Patuh
     Patuh            20          4
     Tidak Patuh      5           12
> Akurasi=100*sum(testing==Prediksi)/NROW(testing)
> Akurasi
[1] 78.04878
> Laju_error=100*(MatriksKonfusi[2]+MatriksKonfusi[3])/NROW(testing)
> Laju_error
[1] 21.95122
```

**Lampiran 7.** *Syntax Naive* Bayes dengan Program R (Lanjutan)

```
#Menggunakan Data Random Keempat
> library(e1071)
> Data= read.table("E://DataWP4.txt",header=TRUE)
> Data$Y<-factor(Data$Y)
> Data$X1<-factor(Data$X1)
> Data$X2<-factor(Data$X2)
> Data$X3<-factor(Data$X3)
> DataTraining<-Data[1:164,]
> DataTesting<-Data[165:205,]
> testing=Data[165:205,1]
> NB<-naiveBayes(Y~X1+X2+X3,data=DataTraining)
> NB
Naive Bayes Classifier for Discrete Predictors
Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)
A-priori probabilities:
Y
   Patuh      TidakPatuh
0.5060976    0.4939024
Conditional probabilities:
                                        X1
    Y              <100 Jt     100-500 Jt   500Jt-1M       1-10M         >10M
  Patuh          0.42168675 0.09638554 0.10843373 0.22891566 0.14457831
  Tidak Patuh   0.37037037 0.19753086 0.09876543 0.24691358 0.08641975
                                        X2
    Y                PT            CV    BUMN/BUMD      KOP       Lainnya
  Patuh          0.43373494 0.43373494 0.00000000 0.03614458 0.09638554
  Tidak Patuh   0.33333333 0.37037037 0.08641975 0.07407407 0.13580247
                                        X3
    Y          Tepat Waktu  Tidak Tepat Waktu
  Patuh          0.7831325 0.2168675
 Tidak Patuh    0.6543210 0.3456790

> Prediksi<-predict(NB,DataTesting,type=c("class"))
> Prediksi
[1] Patuh          Patuh          Tidak Patuh  Patuh          Tidak Patuh  Tidak Patuh
[7] Tidak Patuh   Tidak Patuh   Patuh          Patuh          Tidak Patuh  Tidak Patuh
[13] Patuh         Tidak Patuh   Patuh         Tidak Patuh   Tidak Patuh   Patuh
[19] Tidak Patuh   Patuh          Patuh         Tidak Patuh   Patuh          Patuh
[25] Tidak Patuh   Patuh         Tidak Patuh   Tidak Patuh   Tidak Patuh   Patuh
[31] Tidak Patuh   Tidak Patuh Tidak Patuh   Patuh          Patuh         Tidak Patuh
[37] Tidak Patuh   Tidak Patuh Tidak Patuh   Tidak Patuh   Patuh
Levels: Patuh  Tidak Patuh
```

**Lampiran 7.** *Syntax Naive* Bayes dengan Program R (Lanjutan)

```
> MatriksKonfusi=table(testing,Prediksi)
> MatriksKonfusi
                        Prediksi
testing              Patuh  Tidak  Patuh
     Patuh            15          5
     Tidak  Patuh      2          19
> Akurasi=100*sum(testing==Prediksi)/NROW(testing)
> Akurasi
[1] 82.92683
> Laju_error=100*(MatriksKonfusi[2]+MatriksKonfusi[3])/NROW(testing)
> Laju_error
[1] 17.07317


#Menggunakan  Data Random Kelima
> library(e1071)
> Data= read.table("E://DataWP5.txt",header=TRUE)
> Data$Y<-factor(Data$Y)
> Data$X1<-factor(Data$X1)
> Data$X2<-factor(Data$X2)
> Data$X3<-factor(Data$X3)
> DataTraining<-Data[1:164,]
> DataTesting<-Data[165:205,]
> testing=Data[165:205,1]
> NB<-naiveBayes(Y~X1+X2+X3,data=DataTraining)
> NB
Naive Bayes Classifier  for Discrete  Predictors
Call:
naiveBayes.default(x  = X, y = Y, laplace  = laplace)
A-priori probabilities:
Y
   Patuh       TidakPatuh
     0.5            0.5
Conditional  probabilities:
                                          X1
    Y                <100 Jt     100-500 Jt   500Jt-1M       1-10M          >10M
  Patuh           0.39024390 0.08536585 0.14634146 0.25609756 0.12195122
  Tidak  Patuh   0.32926829 0.19512195 0.09756098 0.29268293 0.08536585
                                          X2
    Y                  PT            CV    BUMN/BUMD      KOP       Lainnya
  Patuh           0.42682927 0.45121951 0.00000000 0.03658537 0.08536585
  Tidak  Patuh   0.31707317 0.36585366 0.08536585 0.12195122 0.10975610
```

**Lampiran 7.** *Syntax Naive* Bayes dengan Program R (Lanjutan)

```
                                        X3
    Y          Tepat Waktu  Tidak Tepat Waktu
  Patuh          0.7682927 0.2317073
  Tidak Patuh   0.6463415 0.3536585

> Prediksi<-predict(NB,DataTesting,type=c("class"))
> Prediksi
 [1] Patuh          Tidak Patuh   Patuh         Patuh          Patuh         Tidak Patuh
[7] Tidak Patuh   Patuh          Tidak Patuh   Patuh          Tidak Patuh   Patuh
[13] Patuh         Tidak Patuh   Patuh         Tidak Patuh   Tidak Patuh   Patuh
[19] Tidak Patuh   Patuh          Patuh         Patuh          Tidak Patuh   Tidak Patuh
[25] Patuh         Tidak Patuh   Patuh         Tidak Patuh   Patuh          Tidak Patuh
[31] Patuh         Tidak Patuh   Tidak Patuh Patuh          Patuh          Patuh
[37] Tidak Patuh   Patuh          Patuh         Patuh          Patuh
Levels: Patuh  Tidak Patuh

> MatriksKonfusi=table(testing,Prediksi)
> MatriksKonfusi
                        Prediksi
testing              Patuh  Tidak Patuh
     Patuh            18         3
     Tidak Patuh      5          15
> Akurasi=100*sum(testing==Prediksi)/NROW(testing)
> Akurasi
[1] 80.4878
> Laju_error=100*(MatriksKonfusi[2]+MatriksKonfusi[3])/NROW(testing)
> Laju_error
[1] 19.5122
```

**Lampiran 8.** Nilai *Posterior* untuk Setiap Kelompok dan Hasil Klasifikasi dengan Metode *Naive* Bayes

| No | Kelompok 1 (Patuh) | Kelompok 2 (Tidak Patuh) | Hasil Klasifikasi |
|---|---|---|---|
| 1 | 0,03935 | 0,02660 | Patuh |
| 2 | 0,03935 | 0,02660 | Patuh |
| 3 | 0,02007 | 0,02342 | Tidak Patuh |
| 4 | 0,02485 | 0,00931 | Patuh |
| 5 | 0,02007 | 0,02108 | Tidak Patuh |
| 6 | 0,01611 | 0,01625 | Tidak Patuh |
| 7 | 0 | 0,01034 | Tidak Patuh |
| 8 | 0 | 0,01034 | Tidak Patuh |
| 9 | 0,07249 | 0,04433 | Patuh |
| 10 | 0,02485 | 0,00931 | Patuh |
| 11 | 0,00459 | 0,01124 | Tidak Patuh |
| 12 | 0,00516 | 0,00625 | Tidak Patuh |
| 13 | 0,07249 | 0,04433 | Patuh |
| 14 | 0,01090 | 0,01561 | Tidak Patuh |
| 15 | 0,07249 | 0,03990 | Patuh |
| 16 | 0,00091 | 0,00312 | Tidak Patuh |
| 17 | 0,01657 | 0,02364 | Tidak Patuh |
| 18 | 0,07249 | 0,03990 | Patuh |
| 19 | 0,01090 | 0,01561 | Tidak Patuh |
| 20 | 0,01864 | 0,01182 | Patuh |
| 21 | 0,03935 | 0,02955 | Patuh |
| 22 | 0,02007 | 0,02342 | Tidak Patuh |
| 23 | 0,01864 | 0,01182 | Patuh |
| 24 | 0,02485 | 0,00931 | Patuh |
| 25 | 0,00138 | 0,00473 | Tidak Patuh |
| 26 | 0,07249 | 0,04433 | Patuh |
| 27 | 0,00091 | 0,00312 | Tidak Patuh |
| 28 | 0,01657 | 0,02364 | Tidak Patuh |
| 29 | 0,01657 | 0,02364 | Tidak Patuh |
| 30 | 0,02485 | 0,01034 | Patuh |
| 31 | 0,00057 | 0,00109 | Tidak Patuh |
| 32 | 0,00328 | 0,00591 | Tidak Patuh |
| 33 | 0,01090 | 0,01405 | Tidak Patuh |
| 34 | 0,01864 | 0,01182 | Patuh |
| 35 | 0,07249 | 0,04433 | Patuh |
| 36 | 0,00057 | 0,00109 | Tidak Patuh |

**Lampiran 8.** Nilai *Posterior* untuk Setiap Kelompok dan Hasil Klasifikasi dengan Metode *Naive* Bayes (Lanjutan)

| No | Kelompok 1 | Kelompok 2 | Hasil Klasifikasi |
|---|---|---|---|
| 37 | 0,01657 | 0,02364 | Tidak Patuh |
| 38 | 0,01090 | 0,11306 | Tidak Patuh |
| 39 | 0,02007 | 0,02342 | Tidak Patuh |
| 40 | 0,02007 | 0,02342 | Tidak Patuh |
| 41 | 0,07249 | 0,04433 | Patuh |

**Lampiran 9.** *Syntax K-Nearest Neighbor* dengan Program R

```
#Menggunakan K=1
library(class)
> Data= read.table("E://DataWP4.txt",header=TRUE)
> str(Data)
'data.frame':    205 obs. of  4 variables:
 $ Y : Factor w/ 2 levels "Patuh",...:  1 1 2 2 ...
 $ X1: int  1 4 1 1 1 4 3 1 2 1 ...
 $ X2: int  2 2 3 3 3 1 2 1 1 2 ...
 $ X3: int  1 2 1 1 1 1 1 1 1 2 ...

> Training=Data[1:164,-1]
> Testing=Data[165:205,-1]
> TrainingTarget=Data[1:164,1]
> TestingTarget=Data[165:205,1]
> Prediksi =knn(Training,Testing,TrainingTarget,k=1)
> Prediksi
[1] Tidak Patuh  Tidak Patuh  Tidak Patuh  Patuh       Tidak Patuh  Patuh
[7] Tidak Patuh  Tidak Patuh  Patuh       Patuh       Tidak Patuh  Patuh
[13] Patuh       Patuh       Patuh       Tidak Patuh Tidak Patuh  Patuh
[19] Patuh       Patuh       Tidak Patuh Tidak Patuh Patuh       Patuh
[25] Tidak Patuh Patuh       Tidak Patuh Tidak Patuh Tidak Patuh  Patuh
[31] Tidak Patuh Tidak Patuh Tidak Patuh Patuh       Patuh       Tidak Patuh
[37] Tidak Patuh Tidak Patuh Tidak Patuh Tidak Patuh Patuh
Levels: Patuh  Tidak Patuh

> MatriksKonfusi=table(TestingTarget,Prediksi)
> MatriksKonfusi
                    Prediksi
TestingTarget  Patuh  Tidak Patuh
    Patuh       15        5
    Tidak Patuh  3        18
> Akurasi=100*sum(TestingTarget==Prediksi)/NROW(TestingTarget)
> Akurasi
[1] 80.4878
>Laju_error=100*(MatriksKonfusi[2]+MatriksKonfusi[3])/NROW(TestingTarget)
> Laju_error
[1] 19.5122
```

**Lampiran 9.** *Syntax K-Nearest Neighbor* dengan Program R (Lanjutan)

```
#Menggunakan K=3
library(class)
> Data= read.table("E://DataWP4.txt",header=TRUE)
> str(Data)
'data.frame':    205 obs. of  4 variables:
 $ Y : Factor w/ 2 levels "Patuh",...:  1 1 2 2 ...
 $ X1: int  1 4 1 1 1 4 3 1 2 1 ...
 $ X2: int  2 2 3 3 3 1 2 1 1 2 ...
 $ X3: int  1 2 1 1 1 1 1 1 1 2 ...

> Training=Data[1:164,-1]
> Testing=Data[165:205,-1]
> TrainingTarget=Data[1:164,1]
> TestingTarget=Data[165:205,1]
> Prediksi =knn(Training,Testing,TrainingTarget,k=3)
> Prediksi
[1] Tidak Patuh  Tidak Patuh  Tidak Patuh  Patuh       Tidak Patuh  Patuh
[7] Tidak Patuh  Tidak Patuh  Patuh       Patuh       Tidak Patuh  Patuh
[13] Patuh       Patuh       Patuh       Patuh       Tidak Patuh  Patuh
[19] Patuh       Patuh       Tidak Patuh Tidak Patuh Patuh        Patuh
[25] Tidak Patuh Patuh       Patuh       Tidak Patuh Tidak Patuh  Patuh
[31] Tidak Patuh Patuh       Tidak Patuh Patuh       Patuh        Tidak Patuh
[37] Tidak Patuh Tidak Patuh Tidak Patuh Tidak Patuh Patuh
Levels: Patuh  Tidak Patuh

> MatriksKonfusi=table(TestingTarget,Prediksi)
> MatriksKonfusi
                    Prediksi
TestingTarget  Patuh  Tidak Patuh
    Patuh        16        4
    Tidak Patuh   5        16
> Akurasi=100*sum(TestingTarget==Prediksi)/NROW(TestingTarget)
> Akurasi
[1] 78.04878
>Laju_error=100*(MatriksKonfusi[2]+MatriksKonfusi[3])/NROW(TestingTarget)
> Laju_error
[1] 21.95122
```

**Lampiran 9.** *Syntax K-Nearest Neighbor* dengan Program R (Lanjutan)

```
#Menggunakan K=5
library(class)
> Data= read.table("E://DataWP4.txt",header=TRUE)
> str(Data)
'data.frame':    205 obs. of  4 variables:
 $ Y : Factor w/ 2 levels "Patuh",...: 1 1 2 2 ...
 $ X1: int  1 4 1 1 1 4 3 1 2 1 ...
 $ X2: int  2 2 3 3 3 1 2 1 1 2 ...
 $ X3: int  1 2 1 1 1 1 1 1 1 2 ...

> Training=Data[1:164,-1]
> Testing=Data[165:205,-1]
> TrainingTarget=Data[1:164,1]
> TestingTarget=Data[165:205,1]
> Prediksi =knn(Training,Testing,TrainingTarget,k=5)
> Prediksi
2 1 2 1 1 1 2 2 2
[40] 2 1
[1] Tidak Patuh  Tidak Patuh  Tidak Patuh  Patuh  Tidak Patuh  Tidak Patuh
[7] Tidak Patuh  Tidak Patuh  Patuh  Patuh  Tidak Patuh  Patuh
[13] Patuh  Patuh  Patuh  Patuh  Tidak Patuh  Patuh
[19] Patuh  Patuh  Tidak Patuh  Tidak Patuh Patuh  Patuh
[25] Tidak Patuh Patuh  Patuh  Tidak Patuh Tidak Patuh  Patuh
[31] Tidak Patuh Patuh  Tidak Patuh  Patuh  Patuh  Patuh
[37] Tidak Patuh Tidak Patuh  Tidak Patuh  Tidak Patuh Patuh
Levels: Patuh  Tidak Patuh

> MatriksKonfusi=table(TestingTarget,Prediksi)
> MatriksKonfusi
                Prediksi
TestingTarget  Patuh  Tidak Patuh
    Patuh        16        4
    Tidak Patuh   5        16
> Akurasi=100*sum(TestingTarget==Prediksi)/NROW(TestingTarget)
> Akurasi
[1] 78.04878
>Laju_error=100*(MatriksKonfusi[2]+MatriksKonfusi[3])/NROW(TestingTarget)
> Laju_error
[1] 21.95122
```

**Lampiran 9.** *Syntax K-Nearest Neighbor* dengan Program R (Lanjutan)

```
#Menggunakan K=7
library(class)
> Data= read.table("E://DataWP4.txt",header=TRUE)
> str(Data)
'data.frame':    205 obs. of  4 variables:
 $ Y : Factor w/ 2 levels "Patuh",...:  1 1 2 2 ...
 $ X1: int  1 4 1 1 1 4 3 1 2 1 ...
 $ X2: int  2 2 3 3 3 1 2 1 1 2 ...
 $ X3: int  1 2 1 1 1 1 1 1 1 2 ...

> Training=Data[1:164,-1]
> Testing=Data[165:205,-1]
> TrainingTarget=Data[1:164,1]
> TestingTarget=Data[165:205,1]
> Prediksi =knn(Training,Testing,TrainingTarget,k=7)
> Prediksi
 [1] Tidak Patuh  Tidak Patuh   Tidak Patuh  Patuh        Tidak Patuh   Patuh
[7] Tidak Patuh  Tidak Patuh   Patuh        Patuh        Tidak Patuh   Patuh
[13] Patuh       Patuh         Patuh        Patuh        Tidak Patuh   Patuh
[19] Patuh       Patuh         Tidak Patuh  Tidak Patuh Patuh          Patuh
[25] Tidak Patuh Patuh         Patuh        Tidak Patuh Tidak Patuh    Patuh
[31] Patuh       Patuh         Tidak Patuh  Patuh        Patuh         Tidak Patuh
[37] Tidak Patuh Tidak Patuh   Tidak Patuh  Tidak Patuh Patuh
Levels: Patuh  Tidak Patuh

> MatriksKonfusi=table(TestingTarget,Prediksi)
> MatriksKonfusi
                    Prediksi
TestingTarget   Patuh   Tidak Patuh
    Patuh        16          4
    Tidak Patuh   6          15
> Akurasi=100*sum(TestingTarget==Prediksi)/NROW(TestingTarget)
> Akurasi
[1] 75.60976
>Laju_error=100*(MatriksKonfusi[2]+MatriksKonfusi[3])/NROW(TestingTarget)
> Laju_error
[1] 24.39024
```

**Lampiran 9.** *Syntax K-Nearest Neighbor* dengan Program R (Lanjutan)

```
#Menggunakan K=9
library(class)
> Data= read.table("E://DataWP4.txt",header=TRUE)
> str(Data)
'data.frame':    205 obs. of  4 variables:
 $ Y : Factor w/ 2 levels "Patuh",...:  1 1 2 2 ...
 $ X1: int   1 4 1 1 1 4 3 1 2 1 ...
 $ X2: int   2 2 3 3 3 1 2 1 1 2 ...
 $ X3: int   1 2 1 1 1 1 1 1 1 2 ...

> Training=Data[1:164,-1]
> Testing=Data[165:205,-1]
> TrainingTarget=Data[1:164,1]
> TestingTarget=Data[165:205,1]
> Prediksi =knn(Training,Testing,TrainingTarget,k=9)
> Prediksi
 [1] Tidak Patuh  Tidak Patuh   Patuh        Patuh        Patuh        Tidak Patuh
[7] Patuh        Patuh        Patuh        Patuh        Tidak Patuh   Patuh
[13] Patuh       Tidak Patuh  Patuh        Patuh        Tidak Patuh   Patuh
[19] Tidak Patuh Patuh        Patuh        Patuh        Patuh        Patuh
[25] Tidak Patuh Patuh        Patuh        Tidak Patuh  Tidak Patuh   Patuh
[31] Patuh       Patuh        Patuh        Patuh        Patuh        Patuh
[37] Tidak Patuh Tidak Patuh  Patuh        Patuh        Patuh
Levels: Patuh  Tidak Patuh

> MatriksKonfusi=table(TestingTarget,Prediksi)
> MatriksKonfusi
                    Prediksi
TestingTarget   Patuh   Tidak Patuh
    Patuh        18          2
    Tidak Patuh  11         10
> Akurasi=100*sum(TestingTarget==Prediksi)/NROW(TestingTarget)
> Akurasi
[1] 68.29268
>Laju_error=100*(MatriksKonfusi[2]+MatriksKonfusi[3])/NROW(TestingTarget)
> Laju_error
[1] 31.70732
```

**Lampiran 10.** Perhitungan Jarak Euclid antara Data *Training* dan Data *Testing*

| Data *Training* | Data *Testing* | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | . . . | 41 |
| 1 | 3,16228 | 3,16228 | 1 | 4,12311 | 1,41421 | . . . | 1,41421 |
| 2 | 1,41421 | 1,41421 | 3 | 1,73205 | 3,16228 | . . . | 3,16228 |
| 3 | 3,60555 | 3,60555 | 1,41421 | 4,47214 | 2,23607 | . . . | 2,23607 |
| 4 | 3,60555 | 3,60555 | 1,41421 | 4,47214 | 2,23607 | . . . | 2,23607 |
| 5 | 3,60555 | 3,60555 | 1,41421 | 4,47214 | 2,23607 | . . . | 2,23607 |
| 6 | 0 | 0 | 3,31662 | 1 | 3,16228 | . . . | 3,16228 |
| 7 | 1,41421 | 1,41421 | 2,23607 | 2,23607 | 2,44949 | . . . | 2,44949 |
| 8 | 3 | 3 | 1,41421 | 4 | 1 | . . . | 1 |
| 9 | 2 | 2 | 1,73205 | 3 | 1,41421 | . . . | 1,41421 |
| 10 | 3,31662 | 3,31662 | 0 | 4,24264 | 1 | . . . | 1 |
| 11 | 3,31662 | 3,31662 | 0 | 4,24264 | 1 | . . . | 1 |
| 12 | 3,16228 | 3,16228 | 1 | 4,12311 | 0 | . . . | 0 |
| 13 | 1,41421 | 1,41421 | 2,23607 | 2,23607 | 2,44949 | . . . | 2,44949 |
| 14 | 3,16228 | 3,16228 | 1 | 4,12311 | 1,41421 | . . . | 1,41421 |
| 15 | 1,41421 | 1,41421 | 2,23607 | 2,23607 | 2,44949 | . . . | 2,44949 |
| 16 | 3,60555 | 3,60555 | 2,44949 | 4,24264 | 3,31662 | . . . | 3,31662 |
| 17 | 3,60555 | 3,60555 | 2,44949 | 4,24264 | 3,31662 | . . . | 3,31662 |
| 18 | 3 | 3 | 1,41421 | 4 | 1 | . . . | 1 |
| 19 | 5,09902 | 5,09902 | 3 | 5,74456 | 4 | . . . | 4 |
| 20 | 2,23607 | 2,23607 | 1,41421 | 3,16228 | 1,73205 | . . . | 1,73205 |
| 21 | 2,23607 | 2,23607 | 1,41421 | 3,16228 | 1,73205 | . . . | 1,73205 |
| 22 | 3 | 3 | 1,41421 | 4 | 1 | . . . | 1 |
| 23 | 1,41421 | 1,41421 | 2,23607 | 2,23607 | 2,44949 | . . . | 2,44949 |
| 24 | 1 | 1 | 3,16228 | 1,41421 | 3,31662 | . . . | 3,31662 |
| 25 | 3 | 3 | 1,41421 | 4 | 1 | . . . | 1 |
| 26 | 2,23607 | 2,23607 | 1,41421 | 3,16228 | 1,73205 | . . . | 1,73205 |
| 27 | 1 | 1 | 4,24264 | 0 | 4,12311 | . . . | 4,12311 |
| 28 | 1 | 1 | 3,16228 | 1,41421 | 3,31662 | . . . | 3,31662 |
| 29 | 1 | 1 | 4,24264 | 0 | 4,12311 | . . . | 4,12311 |
| 30 | 1,41421 | 1,41421 | 2,23607 | 2,23607 | 2 | . . . | 2 |
| 31 | 3,16228 | 3,16228 | 1 | 4,12311 | 0 | . . . | 0 |
| 32 | 1,41421 | 1,41421 | 4,12311 | 1 | 4 | . . . | 4 |
| 33 | 0 | 0 | 3,31662 | 1 | 3,16228 | . . . | 3,16228 |
| 34 | 3,16228 | 3,16228 | 1 | 4,12311 | 0 | . . . | 0 |
| 35 | 4,47214 | 4,47214 | 3,31662 | 5 | 4,24264 | . . . | 4,24264 |

**Lampiran 10.** Perhitungan Jarak Euclid antara Data *Training* dan Data *Testing* (Lanjutan)

| Data *Training* | Data *Testing* | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | . . . | 41 |
| 36 | 3 | 3 | 1,41421 | 4 | 1 | . . . | 1 |
| 37 | 5 | 5 | 3,16228 | 5,65685 | 4,12311 | . . . | 4,12311 |
| 38 | 3 | 3 | 1,41421 | 4 | 1 | . . . | 1 |
| 39 | 3,31662 | 3,31662 | 0 | 4,24264 | 1 | . . . | 1 |
| 40 | 3,16228 | 3,16228 | 1 | 4,12311 | 1,41421 | . . . | 1,41421 |
| 41 | 2,23607 | 2,23607 | 1,41421 | 3,16228 | 1,73205 | . . . | 1,73205 |
| 42 | 1 | 1 | 3,16228 | 1,41421 | 3,31662 | . . . | 3,31662 |
| 43 | 1,41421 | 1,41421 | 4,12311 | 1 | 4 | . . . | 4 |
| 44 | 1 | 1 | 3,16228 | 1,41421 | 3,31662 | . . . | 3,31662 |
| 45 | 1 | 1 | 2,44949 | 2 | 2,23607 | . . . | 2,23607 |
| 46 | 1,41421 | 1,41421 | 4,12311 | 1 | 4,24264 | . . . | 4,24264 |
| 47 | 2 | 2 | 1,73205 | 3 | 1,41421 | . . . | 1,41421 |
| 48 | 4,3589 | 4,3589 | 2 | 5,09902 | 3 | . . . | 3 |
| 49 | 1 | 1 | 3,16228 | 1,41421 | 3,31662 | . . . | 3,31662 |
| 50 | 3 | 3 | 1,41421 | 4 | 1 | . . . | 1 |
| 51 | 3,31662 | 3,31662 | 0 | 4,24264 | 1 | . . . | 1 |
| 52 | 3,16228 | 3,16228 | 1 | 4,12311 | 0 | . . . | 0 |
| 53 | 2,23607 | 2,23607 | 1,41421 | 3,16228 | 1,73205 | . . . | 1,73205 |
| 54 | 5 | 5 | 3,16228 | 5,65685 | 4,12311 | . . . | 4,12311 |
| 55 | 3 | 3 | 1,41421 | 4 | 1 | . . . | 1 |
| 56 | 1 | 1 | 4,24264 | 0 | 4,12311 | . . . | 4,12311 |
| 57 | 1,41421 | 1,41421 | 2,23607 | 2,23607 | 2,44949 | . . . | 2,44949 |
| 58 | 3,16228 | 3,16228 | 1 | 4,12311 | 1,41421 | . . . | 1,41421 |
| 59 | 3 | 3 | 3,74166 | 3,16228 | 4,3589 | . . . | 4,3589 |
| 60 | 1 | 1 | 4,24264 | 0 | 4,12311 | . . . | 4,12311 |
| 61 | 0 | 0 | 3,31662 | 1 | 3,16228 | . . . | 3,16228 |
| 62 | 3,60555 | 3,60555 | 1,41421 | 4,47214 | 2,23607 | . . . | 2,23607 |
| 63 | 4,3589 | 4,3589 | 2 | 5,09902 | 3 | . . . | 3 |
| 64 | 4,47214 | 4,47214 | 3,31662 | 5 | 4,24264 | . . . | 4,24264 |
| 65 | 1,73205 | 1,73205 | 4 | 1,41421 | 4,12311 | . . . | 4,12311 |
| 66 | 1,41421 | 1,41421 | 4,12311 | 1 | 4 | . . . | 4 |
| 67 | 5,09902 | 5,09902 | 3 | 5,74456 | 4 | . . . | 4 |
| . . . | . . . | . . . | . . . | . . . | . . . | . . . | . . . |
| 164 | 3 | 3 | 1,41421 | 4 | 1 | . . . | 1 |

**Lampiran 11.** Hasil Klasifikasi dengan Metode *K-Nearest Neighbor*

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak | Hasil Klasifikasi |
|---|---|---|---|---|---|
| 1 | Tidak Patuh | 1-10 Miliar | PT | Tepat Waktu | Tidak Patuh |
| 2 | Patuh | 1-10 Miliar | PT | Tepat Waktu | Tidak Patuh |
| 3 | Patuh | < 100 Juta | CV | Tidak Tepat Waktu | Tidak Patuh |
| 4 | Patuh | > 10 Miliar | PT | Tepat Waktu | Patuh |
| 5 | Tidak Patuh | < 100 Juta | PT | Tidak Tepat Waktu | Tidak Patuh |
| 6 | Tidak Patuh | < 100 Juta | Lainnya | Tepat Waktu | Patuh |
| 7 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu | Tidak Patuh |
| 8 | Tidak Patuh | < 100 Juta | BUMN/BUMD | Tepat Waktu | Tidak Patuh |
| 9 | Patuh | < 100 Juta | CV | Tepat Waktu | Patuh |
| 10 | Patuh | > 10 Miliar | PT | Tepat Waktu | Patuh |
| 11 | Tidak Patuh | 100-500 Juta | PT | Tidak Tepat Waktu | Tidak Patuh |
| 12 | Patuh | 500 Juta - 1 Miliar | CV | Tidak Tepat Waktu | Patuh |
| 13 | Patuh | < 100 Juta | CV | Tepat Waktu | Patuh |
| 14 | Tidak Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu | Patuh |
| 15 | Patuh | < 100 Juta | PT | Tepat Waktu | Patuh |
| 16 | Tidak Patuh | 1-10 Miliar | Koperasi | Tidak Tepat Waktu | Tidak Patuh |
| 17 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu | Tidak Patuh |
| 18 | Patuh | < 100 Juta | PT | Tepat Waktu | Patuh |
| 19 | Patuh | 1-10 Miliar | CV | Tidak Tepat Waktu | Patuh |
| 20 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu | Patuh |
| 21 | Patuh | 1-10 Miliar | CV | Tepat Waktu | Tidak Patuh |
| 22 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu | Tidak Patuh |
| 23 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu | Patuh |
| 24 | Tidak Patuh | > 10 Miliar | PT | Tepat Waktu | Patuh |
| 25 | Tidak Patuh | 100-500 Juta | Koperasi | Tepat Waktu | Tidak Patuh |
| 26 | Patuh | < 100 Juta | CV | Tepat Waktu | Patuh |
| 27 | Tidak Patuh | 1-10 Miliar | Koperasi | Tidak Tepat Waktu | Tidak Patuh |
| 28 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu | Tidak Patuh |
| 29 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu | Tidak Patuh |
| 30 | Patuh | > 10 Miliar | CV | Tepat Waktu | Patuh |
| 31 | Tidak Patuh | > 10 Miliar | Koperasi | Tidak Tepat Waktu | Tidak Patuh |
| 32 | Patuh | 1-10 Miliar | Koperasi | Tepat Waktu | Tidak Patuh |
| 33 | Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu | Tidak Patuh |
| 34 | Patuh | 500 Juta - 1 Miliar | CV | Tepat Waktu | Patuh |
| 35 | Patuh | < 100 Juta | CV | Tepat Waktu | Patuh |

**Lampiran 11.** Hasil Klasifikasi dengan Metode *K-Nearest Neighbor* (Lanjutan)

| No | Status Pembayaran PPN | Pendapatan (Rupiah) | Bentuk Badan | Status Pelaporan Pajak | Hasil Klasifikasi |
|---|---|---|---|---|---|
| 36 | Tidak Patuh | > 10 Miliar | Koperasi | Tidak Tepat Waktu | Tidak Patuh |
| 37 | Tidak Patuh | 100-500 Juta | CV | Tepat Waktu | Tidak Patuh |
| 38 | Tidak Patuh | 1-10 Miliar | PT | Tidak Tepat Waktu | Tidak Patuh |
| 39 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu | Tidak Patuh |
| 40 | Tidak Patuh | < 100 Juta | CV | Tidak Tepat Waktu | Tidak Patuh |
| 41 | Patuh | < 100 Juta | CV | Tepat Waktu | Patuh |

## RIWAYAT HIDUP

Fatihah Noor Rahmaulidyah lahir pada tanggal 4 Juli 1998 di Kota Samarinda, Kalimantan Timur, merupakan anak bungsu dari tiga bersaudara dari pasangan Bapak Muhammad Noor dan Ibu Ellinah. Memulai pendidikan di Taman Kanak-Kanak Islam Tunas Kartini pada Tahun 2002 kemudian melanjutkan pendidikan sekolah dasar pada Tahun 2004 di Sekolah Dasar Islam Terpadu Cordova. Setelah lulus dari sekolah dasar pada Tahun 2010 melanjutkan pendidikan di Sekolah Menengah Pertama Negeri 1 Samarinda, kemudian melanjutkan ke Sekolah Menengah Atas Negeri 3 Samarinda dan lulus pada Tahun 2016. Pendidikan di perguruan tinggi dimulai pada Tahun 2016 di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman Jurusan Matematika Program Studi Statistika melalui Jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN). Selama menempuh perkuliahan telah aktif dalam organisasi Himpunan Mahasiswa Statistika Universitas Mulawarman, yakni pada periode 2018 dan 2019 serta organisasi Generasi Baru Indonesia periode 2018 dan 2019. Aktif dalam kegiatan-kegiatan akademik seperti menjadi asisten praktikum di Jurusan Matematika untuk beberapa mata kuliah wajib dan pilihan, berpartisipasi dalam lomba 3 *Minutes Thesis* pada Pekan Raya Perpajakan Nasional yang diselenggarakan oleh Politeknik Keuangan Negara STAN. Pada Tahun 2019 mengikuti program Kerja Kuliah Nyata (KKN) Reguler Angkatan XLV di Kelurahan Gunung Elai, Kecamatan Bontang Utara, Kota Bontang selama kurang lebih 50 hari. Kemudian pada tanggal 16 September sampai dengan 8 November 2019 melaksanakan Praktek Kerja Lapangan (PKL) di Kantor Pelayanan Pajak Pratama Samarinda Ulu.