p-ISSN 2085-7829
e-ISSN 2798-3455

Perbandingan Metode Klasifikasi *Naive* Bayes dan *K-Nearest Neighbor* pada Data Status Pembayaran Pajak Pertambahan Nilai di Kantor Pelayanan Pajak Pratama Samarinda Ulu

The Comparison of The Naive Bayes and K-Nearest Neighbor Classification Methods on The Data Payment Status of Value Added Tax at The Samarinda Ulu Pratama Tax Service Office

Fatihah Noor Rahmaulidyah[1], Memi Nor Hayati[2], Rito Goejantoro[3]
[1,3]Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman
[2]Laboratorium Statistika Terapan FMIPA Universitas Mulawarman
E-mail: [1]fn.rahmaulidyah@gmail.com [2]meminorhayati@fmipa.unmul.ac.id

Abstract

Classification is a systematic grouping of objects into certain groups based on the same characteristics. The classification method used in this research are naive Bayes and K-Nearest Neighbor which has a relatively high degree of accuracy. This research aims 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 reported status. Measurement of accuracy using APER in the Naive Bayes method is 17.07% and in K-Nearest Neighbor method is 19,51%. The comparison results of accuracy measurements between the two methods show that the naive Bayes method has a higher level of accuracy than the K-Nearest Neighbor method.

Keywords*: classification, naive Bayes, K-Nearest Neighbor, tax.*

Pendahuluan

Dalam menjalankan roda pemerintahan, negara berupaya untuk menyediakan dan memenuhi segala kebutuhan rakyat melalui pembangunan nasional. Pembangunan nasional dilakukan secara berkesinambungan untuk meningkatkan kesejahteraan masyarakat. Pembiayaan pembangunan yang memerlukan dana besar diupayakan mampu bertumpu pada kemandirian negara. Salah satu sumber dana pembangunan dalam negeri adalah pajak.

Pajak merupakan penghasilan negara yang tujuan pembayarannya digunakan untuk membiayai pembangunan yang berguna bagi kepentingan bersama. Bagi negara pajak merupakan sumber pendapatan atau instrumen yang primer dan strategis. 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). Sebagai salah satu sumber pemasukan negara, PPN lebih menonjol jika dibandingkan dengan PPh. Hal tersebut dikarenakan hampir seluruh transaksi di bidang perdagangan, industri dan jasa pada prinsipnya terkena PPN (Mardiasmo, 2005).

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. Salah satu analisis statistika untuk klasifikasi adalah metode *naive* Bayes. Klasifikasi *naive* Bayes merupakan suatu metode klasifikasi yang berakar pada teorema Bayes yaitu dengan memprediksi peluang di masa depan berdasarkan masa lalu (Bustami, 2013).

Selain metode *naive* Bayes, klasifikasi dapat pula dilakukan dengan menggunakan metode *K-Nearest Neighbor* (K-NN). K-NN termasuk dalam kategori *lazy learner* yang berdasar pada perbandingan *K* tetangga terdekat. Algoritma K-NN merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi atau jarak suatu data dengan data yang lain (Prasetyo, 2012).

Berdasarkan latar belakang tersebut, maka penulis tertarik untuk melakukan penelitian untuk mengetahui perbandingan metode klasifikasi *naive* Bayes dan K-NN pada data status pembayaran pajak ertambahan nilai

p-ISSN 2085-7829
e-ISSN 2798-3455

di kantor pelayanan pajak pratama Samarinda Ulu.

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 statistika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial yang tersimpan dalam *database* besar (Turban dkk, 2005).

Data mining merupakan 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 (Vulandari, 2017). *Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas atau pekerjaan yang dapat dilakukan yaitu deskripsi, estimasi, prediksi, klastering, asosiasi, dan klasifikasi (Larose, 2005).

Klasifikasi

Klasifikasi adalah pengelompokan yang sistematis dari objek, gagasan, buku dan benda-benda lain ke dalam kelompok atau golongan tertentu berdasarkan ciri-ciri yang sama. 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 objek data lain agar diketahui di kelompok mana objek data tersebut dalam model yang sudah disimpannya (Prasetyo, 2014). Klasifikasi disebut juga sebagai metode pengalokasian dengan tujuan untuk memilih atau memasukkan pengamatan (objek baru) ke dalam kelompok yang telah mempunyai label kelompok (Johnson dan Wichern, 2007).

***Naive* Bayes**

Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan 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).

Keuntungan penggunaan metode *naive* Bayes adalah metode ini hanya membutuhkan jumlah data *training* yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian (Santosa, 2007).

Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (1)$$

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

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.

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

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

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, $i = 1,2, \dots, n$
x_{jk}^* = nilai variabel bebas ke-k dari data *testing* ke-j, $j = 1,2, \dots, n^*$
p = banyaknya variabel bebas

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 (Johnson dan Wichern, 2007). Menghitung nilai APER dapat dilakukan melalui tabel 1.

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*)

Tabel 1. 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 objek dari c_1 tepat diklasifikasikan sebagai $\hat{c}_1$

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

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

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

dengan perhitungan nilai APER sebagai berikut:

$$APER = \frac{\text{jumlah obyek yang salah klasifikasi}}{\text{jumlah prediksi yang dilakukan}} \times 100\% = \frac{f_{12}+f_{12}}{f_{12}+f_{12}+f_{12}+f_{12}} \times 100\% \quad (3)$$

Metode Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Kantor Pelayanan Pajak Pratama Samarinda Ulu. Adapun data yang digunakan adalah data Wajib Pajak Badan mengenai status pembayaran pajak, pendapatan, bentuk badan, dan status pelaporan pajak.

Langkah-langkah yang dilakukan dalam pengklasifikasian dengan metode *naive* Bayes adalah sebagai berikut :

1. Menghitung nilai probabilitas awal (*prior*).
2. Menghitung probabilitas setiap variabel bebas pada setiap kelompok.
3. Menghitung probabilitas akhir (*posterior*) pada masing-masing kelas.
4. Mencari nilai maksimum pada kedua kelompok dan menentukan hasil klasifikasi objek.
5. Mengevaluasi hasil klasifikasi dengan menghitung nilai APER.

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

1. Menentukan nilai parameter K, pada penelitian ini menggunakan $K = 1,3,5,7,9$
2. Menghitung jarak Euclid objek terhadap data *training*.
3. Mengurutkan data yang mempunyai jarak terkecil hingga terbesar.
4. Menentukan hasil klasifikasi dengan menggunakan kategori *nearest neighbor* yang paling banyak.
5. Mengevaluasi hasil klasifikasi dengan menghitung nilai APER.

Setelah diperoleh nilai APER pada kedua metode tersebut, selanjutnya dilakukan perbandingan hasil klasifikasi. Ketepatan klasifikasi dengan nilai yang terbesar akan diambil untuk kemudian ditarik kesimpulan.

Hasil dan Pembahasan

1. Klasifikasi *naive* Bayes

Pada proses klasifikasi metode *naive* Bayes menggunakan data *training* 80% sebanyak 164 data, sedangkan data *testing* 20% sebanyak 41 data dengan pengulangan sebanyak 5 kali dan menghasilkan akurasi terbaik yaitu dengan laju error sebesar 17,07%, dimana pengklasifikasian-nya dapat dilihat pada Tabel 2.

2. Klasifikasi *K-Nearest Neighbor*

Pada proses klasifikasi metode *naive* Bayes menggunakan data *training* 80% sebanyak 164 data, sedangkan data *testing* 20%

p-ISSN 2085-7829
e-ISSN 2798-3455

sebanyak 41 data, serta menggunakan $K = 1,3,5,7,9$ dengan pengulangan sebanyak 5 kali dan menghasilkan laju error seperti pada Tabel 3.

Tabel 2. 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

Tabel 3. 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 3 dapat diketahui bahwa pada percobaan menggunakan 1-NN pada data randomisasi ke-4 mendapatkan nilai APER yang paling kecil yaitu 19,51%, 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, dimana pengklasifikasiannya dapat dilihat pada Tabel 4.

Tabel 4. 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

Perbandingan Tingkat Akurasi Klasifikasi

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 5. Berdasarkan Tabel 5 dapat diketahui bahwa pengklasifikasian dengan metode *naive* Bayes memperoleh nilai APER yang lebih kecil dibandingkan dengan metode K-NN. Hal ini menunjukkan bahwa metode *naive* Bayes memberikan ketepatan prediksi klasifikasi yang lebih baik.

Tabel 5. Perbandingan Tingkat Akurasi Klasifikasi

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

Kesimpulan

Berdasarkan hasil analisis data dan pembahasan 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.

Daftar Pustaka

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

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

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.

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.

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

Turban, E., Aronson, J. E., dan Liang, T. P. (2005). *Decision Support Systems and Intelligent Systems Edisi Bahasa Indonesia Jilid 1*. 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.

p-ISSN 2085-7829
e-ISSN 2798-3455