SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika

https://ejurnal.itats.ac.id/snestik dan https://snestik.itats.ac.id

Informasi Pelaksanaan :
SNESTIK IV – Surabaya, 27 April 2024
Ruang Seminar Gedung A, Kampus Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:
DOI : 10.31284/p.snestik.2024.5880

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

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

Veronica Nadya Agatha

Universitas Sebelas Maret

e-mail: veronicanadyaa@student.uns.ac.id

ABSTRACT

Value Added Tax (VAT) is a tax added to a transaction. VAT also serves as a source of state revenue to support national development. This research was motivated by the city of Surakarta which had obtained the title as the second tax-abiding city in Central Java, so it is necessary to review the suitability between the implementation, deposit, and reporting of VAT with the regulations applicable to the claim. This study uses tax payment compliance data based on data from the Surakarta Pratama Tax Service Office with a tax payment status classification label. In this study, a comparison of the naïve Bayes and K-Nearest Neighbor methods was carried out in the classification of VAT payment status data in Surakarta, 2023. The comparison has significant urgency in helping the government or tax agency to optimize tax collection, formulate more effective policies, improve tax compliance in the community, and improve overall administrative efficiency. To evaluate accurate results, accuracy level measurements are carried out using Apparent Error Rate (APER). The results of this study showed that the naïve Bayes method had better classification performance with an APER value of 8.31% while the K-NN method was 13.07%.

Keywords: *Value Added Tax, Classification, naïve Bayes, K-NN, APER*

ABSTRAK

Pajak Pertambahan Nilai (PPN) merupakan pajak yang ditambahkan atas suatu transaksi. PPN juga berfungsi sebagai sumber penerimaan negara untuk menyokong pembangunan nasional. Penelitian ini

dilatarbelakangi oleh kota Surakarta yang sempat memperoleh predikat sebagai kota kedua taat pajak di Jawa Tengah sehingga perlu dilakukan peninjauan kesesuaian antara pelaksanaan, penyetoran, dan pelaporan PPN dengan peraturan yang berlaku atas klaim tersebut. Penelitian ini menggunakan data kepatuhan pembayaran pajak berdasarkan data dari Kantor Pelayanan Pajak Pratama Surakarta dengan label klasifikasi status pembayaran pajak. Pada penelitian ini, dilakukan perbandingan metode *naive* Bayes dan *K-Nearest Neighbor* pada klasifikasi data status pembayaran PPN di Surakarta tahun 2023. Perbandingan yang dilakukan memiliki urgensi yang signifikan dalam membantu pemerintah atau lembaga pajak untuk mengoptimasi pengumpulan pajak, menyusun kebijakan yang lebih efektif, meningkatkan ketaatan pajak di masyarakat, dan meningkatkan efisiensi administrasi secara keseluruhan. Untuk mengevaluasi hasil yang akurat, dilakukan pengukuran tingkat akurasi menggunakan *Apparent Error Rate* (APER). Hasil penelitian ini menunjukkan bahwa metode *naive* Bayes memiliki performa klasifikasi yang lebih baik dengan nilai APER sebesar 8,31% sedangkan metode K-NN sebesar 13,07%.

Kata Kunci: Pajak Pertambahan Nilai, Klasifikasi, *naïve* Bayes, K-NN, APER

PENDAHULUAN

Pajak Pertambahan Nilai (PPN) merupakan pajak yang ditambahkan atas suatu transaksi. PPN sangat penting untuk meningkatkan penerimaan negara. Hal ini dikarenakan PPN akan dikenakan pada setiap barang dan kebutuhan masyarakat pada berbagai bidang hasil produksi. Di Indonesia, pemerintah menerapkan tiga sistem pemungutan pajak yaitu *official assesment system, self assesment system,* dan *withholding system*. Pada penerapannya, *self assesment system* dilakukan pada pemungutan pajak penghasilan, pajak penjualan atas barang mewah, sebagian pada pajak bumi dan bangunan, serta pajak pertambahan nilai. Sistem ini berarti memberikan kepercayaan kepada wajib pajak untuk memenuhi dan melaksanakan sendiri kewajiban dan hak perpajakannya, sehingga wajib pajak melakukan sendiri perhitungan pajak terutangnya, menyetorkan ke bank persepsi, dan kemudian melaporkan secara teratur ke Kantor Pelayanan Pajak dalam bentuk surat pemberitahuan. Namun, hal ini tidak berjalan lancar dikarenakan Direktorat Jenderal Pajak belum memiliki perangkat pengawasan yang memadai seperti data lengkap dan akurat mengenai usaha Wajib Pajak. Selain itu, banyak Wajib Pajak yang kurang berkenan mendaftarkan dirinya. Oleh karena itu, penting pada suatu sistem perpajakan untuk memantau status kepatuhan pembayaran secara efektif dan efisien.

Klasifikasi adalah sebuah proses penemuan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan untuk memprediksi kelas yang label kelasnya tidak diketahui. Metode *machine learning* menjadi salah satu pendekatan dalam tahapan proses klasifikasi yang dapat diterapkan untuk memantau status kepatuhan masyarakat pada pembayaran PPN dengan syarat pemilihan metodenya harus tepat sehingga prediksi yang dihasilkan tidak akan keliru. Penelitian ini dilatarbelakangi oleh kota Surakarta yang sempat memperoleh predikat sebagai kota kedua taat pajak di Jawa Tengah sehingga perlu dilakukan peninjauan kesesuaian antara pelaksanaan, penyetoran, dan pelaporan PPN dengan peraturan yang berlaku atas klaim tersebut. Penelitian ini menggunakan data kepatuhan pembayaran pajak berdasarkan data dari Kantor Pelayanan Pajak Pratama Surakarta dengan label klasifikasi status pembayaran pajak.

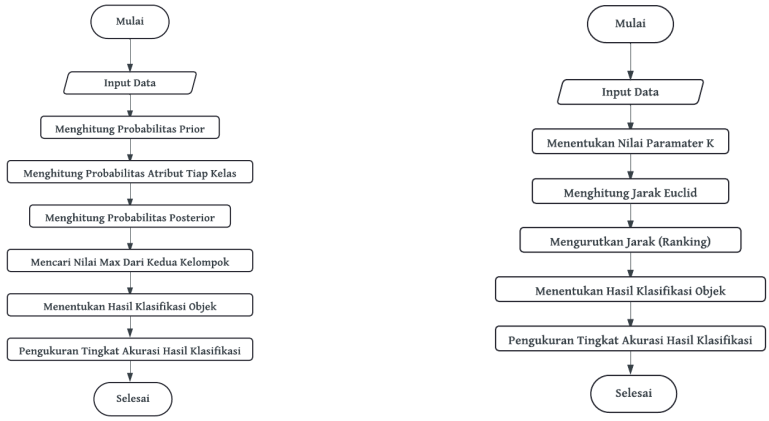
Pada penelitian ini, dilakukan perbandingan metode *naive* Bayes dan *K-Nearest Neighbor* pada klasifikasi data status pembayaran PPN di Surakarta tahun 2023. Kedua metode memiliki tingkat akurasi yang cukup tinggi. Hal ini dapat dilihat dari penelitian-penelitian terdahulu, diantaranya (Wisdayani, 2019) yang memperoleh nilai akurasi sebesar 88,82% untuk klasifikasi tingkat kecelakaan lalu lintas di Pati, Jawa Tengah. Adapun penelitian oleh Anam dkk. (2021) yang menerapkan *naive* Bayes, K-NN dan *decision tree* untuk menganalisis sentimen pada interaksi netizen dan pemerintah yang menuai hasil akurasi 100% untuk metode *naive* Bayes, tingkat akurasi 98,25% untuk K-NN, dan metode *decision tree* sebesar 62,28%.

Dari pemaparan diatas, penelitian ini memiliki urgensi yang signifikan dalam membantu pemerintah atau lembaga pajak untuk mengoptimasi pengumpulan pajak, menyusun kebijakan

yang lebih efektif, meningkatkan ketaatan pajak di masyarakat, dan meningkatkan efisiensi administrasi secara keseluruhan di Kantor Pelayanan Pajak Pratama Surakarta.

METODE

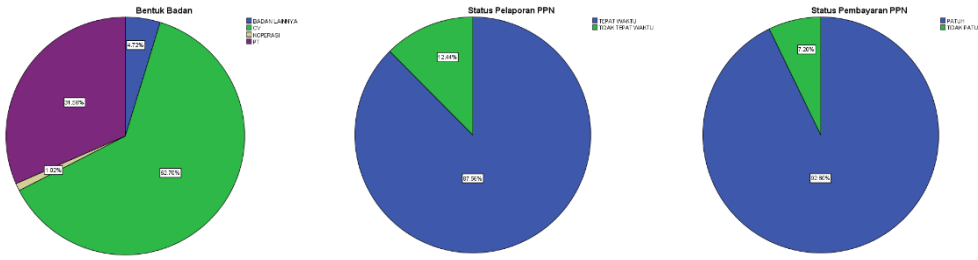
Penelitian ini menggunakan algoritma *naive* Bayes dan *K-Nearest Neighbor* dengan metode evaluasi *matrix confusion* untuk meramalkan data Wajib Pajak menggunakan data status pembayaran Pajak Pertambahan Nilai tahun 2023 dari Kantor Pelayanan Pajak Pratama Surakarta. Hal ini disempurnakan dengan menggunakan *software* R untuk pengolahan datanya. Berikut merupakan tahapan-tahapan untuk memaksimalkan proses penelitian.



(a) (b)

Gambar 1. a) Penggunaan *software* lucidchart untuk pembuatan *flowchart* metode *naive* Bayes, b) Penggunaan *software* lucidchart untuk pembuatan *flowchart* metode K-NN.

Pada data tersebut, terdapat 8177 Wajib Pajak Badan yang terdaftar. Adapun variabel yang diperlukan seperti satu variabel terikat yakni Status Pembayaran PPN (Y) disertai dua variabel bebas yakni Bentuk Badan (X_1) dan Status Pelaporan PPN (X_2). Sebelum memulai tahapan seperti yang tertera pada flowchart, diperlukan analisis statistik deskriptif yang dapat dijabarkan dengan diagram berikut.



(a) (b) (c)

Gambar 2. a) Bentuk Badan (X_1), b) Status Pelaporan PPN (X_2), c) Status Pembayaran PPN (Y)

Sumber: dokumen pribadi

HASIL DAN PEMBAHASAN

Pembahasan Data I

Dilakukan pembagian data menjadi *data training* dan *data testing* dengan komposisi 80% : 20% sehingga dari 8177 data, sebanyak 6494 data teratas berperan sebagai data *training* dan 1683 sisanya berperan sebagai data *testing*. Setelah itu, metode *naïve* Bayes dan K-NN dapat diterapkan.

Tabel 1. Probabilitas Prior Setiap Kelompok

Status Pembayaran PPN	Jumlah	Probabilitas
Patuh	5950	0.9162
Tidak Patuh	544	0.0837

Keterangan: Data yang digunakan untuk mencari probabilitas prior setiap kelompok ialah data training

Tabel 2. Probabilitas Variabel Bentuk Badan Setiap Kelompok (X_1)

Bentuk Badan Instansi	Status Pembayaran Pajak	
	Patuh	Tidak patuh
PT	0.3337	0.2941
CV	0.6255	0.4558
Koperasi	0.0119	0.0018
Badan Lainnya	0.0287	0.2481

Tabel 3. Probabilitas Variabel Status Pelaporan PPN Setiap Kelompok (X_2)

Status Lapor Pajak	Status Pembayaran Pajak	
	Patuh	Tidak patuh
Patuh	0.9490	0.2573
Tidak Patuh	0.051	0.7426

Tabel 4. Matriks Konfusi Metode *naïve* Bayes

Retrieved	*Predicted*		Total
	Patuh	Tidak patuh	
Patuh	1526	112	1638

Tidak Patuh	23	22	45
Total	1549	134	1683

Dari proses dengan metode *naïve* Bayes diatas, diperoleh nilai APER sebesar 3,83%.Untuk tahapan selanjutnya, dilakukan penerapan metode K-NN dengan langkah awal yang paling penting yakni menentukan nilai parameter k. Pada perhitungan contoh digunakan K = 5. Langkah kedua ialah menghitung jarak Euclid antara data *training* dan data *testing* pada setiap observasi data *training.* Perhitungan jarak Euclid pada observasi ini memperoleh hasil 0. Setelah itu, hasil perhitungan jarak Euclid berperan dalam pengurutan *ascending* sehingga menghasilkan data *ranking* sebagai berikut.

Tabel 5. Perhitungan Jarak Euclid dalam Pengurutan Data Ranking

Rank	Data *Training*		Jarak Euclidean
	Sampel	Klasifikasi	
1	25	Patuh	0
2	38	Patuh	0
3	48	Patuh	0
4	50	Patuh	0
5	51	Tidak Patuh	0
6	69	Tidak Patuh	0
7	289	Patuh	0
…	…	…	...
1683	2	Patuh	3

Nilai parameter K=5 sehingga dilihat 5 ranking teratas pada *data training*, diketahui jika observasi yang diklasifikasikan Patuh terdapat 4 observasi dan 1 observasi dikategorikan Tidak Patuh, maka observasi 1 pada *data testing* akan status pembayaran PPN akan diklasifikasikan Patuh. Setelah dihasilkan hasil prediksi pada setiap observasi pada data *testing* sebanyak 1683 data. Hasil klasifikasi metode K-NN juga menggunakan matriks konfusi untuk menghitung nilai APER. Berikut penyajian matriks konfusi hasil metode klasifikasi K-NN.

Tabel 6. Matriks Konfusi dengan K-NN

Retrieved	Predicted		Total
	Patuh	Tidak patuh	
Patuh	1365	8	1373

Tidak Patuh	212	98	310
Total	1577	106	1683

Berdasarkan Tabel 6 tersebut dapat diperoleh nilai APER untuk metode K-NN sebesar 13.07%.

KESIMPULAN

Berdasarkan analisis yang dilakukan pada penelitian ini, diperoleh bahwa nilai akurasi pada kedua metode klasifikasi yang digunakan menggunakan pendekatan nilai APER dapat dikategorikan sebagai pengklasifikasian yang baik dengan nilai APER pada metode *naive* Bayes sebesar 8.31% dan K-NN sebesar 13.07 %. Pada kedua metode tersebut, terlihat jika nilai APER pada pengklasifikasian menggunakan metode *naive* Bayes lebih kecil jika dibandingkan dengan metode K-NN sehingga disimpulkan pada pengklasifikasian Status Pembayaran PPN di KPP Pratama Kota Surakarta tahun 2024 dengan metode *naive* Bayes memberikan ketepatan hasil prediksi klasifikasi yang lebih baik dibandingkan dengan metode K-NN. Perbandingan yang sudah dilakukan memiliki urgensi yang signifikan dalam membantu pemerintah atau lembaga pajak untuk mengoptimasi pengumpulan pajak, menyusun kebijakan yang lebih efektif, meningkatkan ketaatan pajak di masyarakat, dan meningkatkan efisiensi administrasi secara keseluruhan.

DAFTAR PUSTAKA

[1] A. Arham and A. Firmansyah, "THE ROLE OF BEHAVIORAL THEORY IN THE RESEARCH OF MSMEs TAX COMPLIANCE IN INDONESIA," *Riset*, vol. 3, no. 1, pp. 417–432, Mar. 2021, doi: 10.37641/riset.v3i1.71.

[2] A.-D. Pham, N.-T. Ngo, T. T. Ha Truong, N.-T. Huynh, and N.-S. Truong, "Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability," *J Clean Prod*, vol. 260, p. 121082, Jul. 2020, doi: 10.1016/j.jclepro.2020.121082

[3] S. E. Kemp, M. Ng, T. Hollowood, and J. Hort, "Introduction to Descriptive Analysis," in *Descriptive Analysis in Sensory Evaluation*, Wiley, 2018, pp. 1–39. doi: 10.1002/9781118991657.ch1.

[4] F. N. Rahmaulidyah, M. N. Hayati, and R. Goejantoro, "Perbandingan Metode Klasifikasi Naive Bayes dan K-Nearest Neighbor pada Data Status Pembayaran Pajak Pertambahan Nilai di Kantor Pelayanan Pajak Pratama Samarinda Ulu," *EKSPONENSIAL*, vol. 12, no. 2, p. 161, Dec. 2021, doi: 10.30872/eksponensial.v12i2.809.

[5] F.-J. Yang, "An Implementation of Naive Bayes Classifier," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, IEEE, Dec. 2018, pp. 301–306. doi: 10.1109/CSCI46756.2018.00065.

[6] P. Cunningham and S. J. Delany, "k-Nearest Neighbour Classifiers - A Tutorial," *ACM Comput Surv*, vol. 54, no. 6, pp. 1–25, Jul. 2022, doi: 10.1145/3459665.

[7] P.-E. Danielsson, "Euclidean distance mapping," *Computer Graphics and Image Processing*, vol 14, no.3, pp. 227-248, Nov. 1980, doi: 10.1016/0146-664X(80)90054-4

[8] S. Zhang, D. Cheng, Z. Deng, M. Zong, and X. Deng, "A novel k NN algorithm with data-driven k parameter computation," *Pattern Recognit Lett*, vol. 109, pp. 44–54, Jul. 2018, doi: 10.1016/j.patrec.2017.09.036.

[9] M. R. Silvestre and L. L. Ling, "Pruning methods to MLP neural networks considering proportional apparent error rate for classification problems with unbalanced data," *Measurement*, vol. 56, pp. 88–94, Oct. 2014, doi: 10.1016/j.measurement.2014.06.018.