

Kajian Algoritma Jaringan Syaraf Tiruan untuk Mendeteksi Secara Dini Kepatuhan Wajib Pajak Orang Pribadi

Tatang Rohana
Sekolah Tinggi Manajemen Informatika dan Komputer
Cikarang
cctatang@gmail.com

Muh Arifuddin
Direktorat Jendral Pajak Kementerian Keuangan Republik
Indonesia
ariflasis@gmail.com

Abstrak—Wajib Pajak Orang Pribadi diberi kebebasan untuk memenuhi kewajiban perpajakan sesuai dengan peraturan perpajakan yang berlaku. Kepatuhan pemenuhan kewajiban tersebut, salah satunya ditunjukkan dengan pelaporan Surat Pemberitahuan (SPT) Tahunan. Dengan *data mining*, kepatuhan Wajib Pajak Orang Pribadi dapat dideteksi. Wajib Pajak Orang Pribadi akan diklasifikasikan, mana yang lapor dan mana yang tidak lapor. Algoritma berbasis jaringan syaraf tiruan, yaitu *Multilayer Perceptron (MLP)* dan *Support Vector Machine (SVM)* akan diterapkan untuk mengetahui akurasi dari tugas klasifikasi. Penerapan itu selanjutnya diuji dengan *confusion matrix* dan kurva *Receiver Operating Characteristic (ROC)* sehingga dapat diketahui algoritma mana yang memperoleh akurasi yang lebih tinggi.

Kata kunci—Wajib Pajak Orang Pribadi, *Data Mining*, *Klasifikasi*, *Multilayer Perceptron (MLP)*, *Support Vector Machine (SVM)*

I. PENDAHULUAN

Kegiatan perencanaan, evaluasi dan pengambilan keputusan dapat dilakukan dengan baik apabila para pengambil keputusan memiliki informasi yang lengkap, cepat, tepat dan akurat. Hambatan yang sering ditemukan adalah kenyataan bahwa ketika kita sudah memiliki data yang cukup lengkap dan besar akan tetapi kita seringkali mengalami kesulitan dalam menyajikan dan mengolah data tersebut sehingga menjadi informasi yang berguna dan mudah dipahami dengan tinjauan dari berbagai sudut pandang dan tingkat rincian data yang diinginkan. Hambatan lain yang sering juga dihadapi adalah bagaimana melakukan proses penggalian/ekstraksi informasi dari data yang masih tersembunyi untuk selanjutnya diolah menjadi informasi yang terus berkembang menjadi pengetahuan (*knowledge*) yang bermanfaat bagi proses pengambilan keputusan.

Salah satu teknik yang dapat digunakan dalam proses penggalian informasi dari data yang masih tersembunyi dalam jumlah yang besar dan kompleks adalah teknik *data mining*. Output dari penggunaan teknik *data mining* adalah diperolehnya karakteristik atau pola dari data tersebut. Dalam kajian ini, penerapan teknik *data mining* digunakan untuk

mengidentifikasi atau mengenal karakteristik dari kepatuhan wajib pajak orang pribadi. Hal ini dilakukan untuk mengetahui Apakah Wajib Pajak jujur menghitung seluruh penghasilannya? Apakah Wajib Pajak sudah menghitung pajak sesuai dengan aturan yang telah ditetapkan? Apakah Wajib Pajak sudah menyetorkan pajaknya? Apakah Wajib Pajak sudah mengisi Surat Pemberitahuan (SPT) dengan benar? Apakah Wajib Pajak sudah melaporkan Surat Pemberitahuan (SPT) dan yang lain – lainnya. Kajian ini diperlukan karena berdasarkan data pajak tahun 2008 sampai dengan tahun pajak 2011, terjadi penurunan persentasi Wajib Pajak Orang Pribadi Efetif yang melaporkan Surat Pemberitahuan (SPT) Tahunan [2]. Untuk tahun pajak 2011, Wajib Pajak Orang Pribadi Efetif yang mempunyai kewajiban melaporkan Surat Pemberitahuan (SPT) Tahunan ke Kantor Pelayanan Pajak (KPP) Pratama Jakarta Palmerah adalah 46.975. Sedangkan yang melaporkan Surat Pemberitahuan (SPT) Tahunan adalah 17.874. Sehingga persentasinya adalah 38,1% [5]. Hal ini tentu saja dapat digunakan sebagai salah satu indikasi kesadaran wajib pajak pribadi dalam melaksanakan kewajibannya masih kurang.

Untuk meningkatkan kesadaran para wajib pajak pribadi, pertama kali yang harus dilakukan adalah mengenali Wajib Pajak yang terdaftar dalam administrasi perpajakan Kantor Pelayanan Pajak (KPP). Wajib Pajak diklasifikasikan mana yang patuh (lapor Surat Pemberitahuan (SPT) Tahunan) dan mana yang tidak patuh (tidak lapor Surat Pemberitahuan (SPT) Tahunan). Untuk lebih mendapatkan atau menggali data dengan tingkat akurasi data yang lebih baik, dalam implementasi *data mining* ini algoritma yang digunakan adalah jaringan syaraf tiruan. Adapun alasan digunakannya algoritma jaringan syaraf tiruan, karena jaringan syaraf tiruan dapat menurunkan hubungan linier dan *non-linier* yang kompleks, *input/output* yang fleksibel, dan jaringan syaraf tiruan lebih tahan dari pengaruh kebisingan (*noise*) data dibandingkan dengan model yang lainnya.

II. PENELITIAN TERKAIT

Dalam penelitian ini, ada beberapa penelitian yang dilakukan oleh Peneliti sebelumnya yang digunakan sebagai

referensi penelitian. Penelitian tersebut diuraikan sebagai berikut.

A. Penerapan *Particle Swarm Optimization* untuk Seleksi Atribut

Dalam penelitian yang berjudul Penerapan *Particle Swarm Optimization* untuk Seleksi Atribut pada Metode *Support Vector Machine* untuk Prediksi Penyakit Diabetes [9], peneliti melakukan prediksi terjadinya penyakit *diabetes melitus*.

Data penelitian yang diperoleh adalah data sekunder karena diperoleh dari Pima Indian *diabetes database*. Masalah yang harus dipecahkan di sini adalah prediksi terjadinya penyakit *diabetes melitus* dalam waktu 5 tahun dengan menggunakan data Pima yang berisi 786 orang yang diperiksa dan sebanyak 500 pasien tidak terdeteksi terkena penyakit *diabetes melitus*, sehingga 268 pasien terdeteksi penyakit *diabetes melitus*. Atribut penyakit *diabetes melitus* adalah berapa kali hamil, konsentrasi glukosa, tekanan darah, ketebalan lipatan kulit, serum insulin, indeks massa tubuh, fungsi silsilah *diabetes melitus*, dan umur serta kelas sebagai label yang terdiri atas ya dan tidak.

Penelitian ini berusaha membandingkan prediksi terjadinya penyakit *diabetes melitus* antara algoritma *Support Vector Machine* (SVM) dan *Support Vector Machine* (SVM) berbasis *Particle Swarm Optimization* (PSO). Perangkat lunak yang digunakan adalah RapidMiner 5.2.

Penelitian diawali dengan melakukan *data training* terhadap kedua algoritma. Selanjutnya, keduanya dievaluasi dan diuji dengan *confusion matrix* dan kurva *Receiver Operating Characteristic* (ROC). Dari hasil evaluasi, diketahui bahwa akurasi *Support Vector Machine* (SVM) adalah 74,21% dan akurasi *Support Vector Machine* (SVM) berbasis *Particle Swarm Optimization* (PSO) adalah 77,36%. Sehingga algoritma yang terpilih adalah *Support Vector Machine* (SVM) berbasis *Particle Swarm Optimization* (PSO) [9].

B. Komparasi Penerapan Algoritma C 4.5, K-Nearest Neighbor, dan Naive Bayes

Penelitian ini terinspirasi dari informasi bahwa telah terjadi peningkatan kredit macet dari tahun 2010 sampai dengan 2012 pada koperasi yang menjadi objek penelitian. Penelitian ini diberi judul Komparasi Penerapan Algoritma C 4.5, K-Nearest Neighbor, dan Naive Bayes dalam Penentuan Kelayakan Peminjaman Kredit Nasabah Koperasi [14].

Peneliti ingin membandingkan tingkat akurasi dari penerapan algoritma C 4.5, K-Nearest Neighbor, dan Naive Bayes pada data transaksi kredit koperasi mulai tahun 2010 sampai dengan 2012 untuk mengklasifikasikan

kredit yang diberikan lancar atau macet. Perangkat lunak yang digunakan adalah RapidMiner 5.2.

Data yang didapat dari penelitian ini sebanyak 2277 records, baik yang macet ataupun lancar, yang terdiri dari 10 atribut. Atribut yang digunakan adalah jenis kelamin, tingkat umur, pendidikan terakhir, status pernikahan, pekerjaan, penghasilan, jumlah pinjaman, jenis pinjaman, jangka waktu, dan kegunaan pinjaman serta kelas sebagai label yang terdiri atas lancar dan macet.

Setelah dihilangkan duplikasi dan *null value* menjadi 513 records. Untuk *data testing* atau sampel diambil sebanyak 100 records dan untuk penerapan diambil sebanyak 13 records. Untuk pemilihan data tersebut menggunakan tehnik *Systematic Random Sampling*.

Ketiga algoritma tersebut diterapkan dengan menggunakan *data training* berjumlah 400 records dan *data testing* berjumlah 100 records. Hasil penerapan tersebut dievaluasi dan diuji dengan *confusion matrix* dan kurva *Receiver Operating Characteristic* (ROC). Tingkat akurasi yang dihasilkan adalah C 4.5 sebesar 93%, K-Nearest Neighbor sebesar 85,4%, dan Naive Bayes sebesar 89,6% [14]. Oleh karena itu, algoritma yang terpilih adalah C 4.5 dengan akurasi sebesar 93%. Selanjutnya, Sardiarinto menerapkannya ke dalam bahasa pemrograman PHP dan diuji dengan data baru.

C. Kajian Penerapan Model Neural Network untuk Prediksi Penyakit Hati

Dalam penelitian ini, peneliti memprediksi terjadinya penyakit hati. Penelitian ini diberi judul Kajian Penerapan Model Neural Network untuk Prediksi Penyakit Hati [16]. Data yang digunakan dalam penelitian ini bersumber dari alamat web: <http://archive.ics.uci.edu/ml/>.

Data tersebut merupakan hasil pemeriksaan terhadap 583 orang dari wilayah Andhra Pradesh, India yang diperiksa dengan hasil 416 orang terdeteksi menderita penyakit hati dan 167 orang tidak terdeteksi menderita penyakit hati. Parameter *input* dalam jaringan syaraf tiruan ini terdiri dari sepuluh variabel (atribut), yaitu: umur, jenis kelamin, Total Bilirubin ((TB), Direct Bilirubin (DB), Alkaline Phosphatase (Alkphos), Alamine Aminotransferase (Sgpt), Aspartate Aminotransferase (Sgot), Total Proteins (TP), Albumin (ALB), dan Albumin and Globulin Rasio (A/G Rasio). Sedangkan keluaran yang dihasilkan adalah variabel *output* yang bernilai 1 untuk positif penderita penyakit hati dan 2 untuk negatif penderita penyakit hati.

Setiadi membandingkan penerapan metode *artificial neural network* dengan algoritma *Multilayer Perceptron* (MLP) dan *Radial Basis Function* (RBF). Perangkat lunak yang digunakan untuk menganalisa adalah SPSS Neural Networks 17.0.

Langkah awal yang dilakukan oleh peneliti adalah melakukan *preprocessing* terhadap data penelitian. Selanjutnya, diolah dengan menggunakan algoritma *Multilayer Perceptron* (MLP) dan *Radial Basis Function* (RBF). Hasil penerapan tersebut dievaluasi dan diuji dengan *confusion matrix* dan kurva *Receiver Operating Characteristic* (ROC). Perhitungan dengan menggunakan algoritma *Multilayer Perceptron* (MLP) mendapatkan akurasi 0,734568 dan algoritma *Radial Basis Function* (RBF) mendapatkan akurasi 0,0728395. Oleh karena itu, algoritma yang terpilih adalah *Multilayer Perceptron* (MLP) dengan akurasi 0,734568 [16].

III. PEMODELAN DATA MINING

A. Sumber Data

Data yang diolah merupakan data sekunder. Data sekunder yang diperoleh dari objek penelitian adalah sebanyak 46.975 records. Sebelum diolah dengan *data mining*, data tersebut diproses terlebih dulu pada tahapan *data preparation* yang terdiri atas *data cleaning*, *data option*, *data preprocessing*, dan *data expression*. Proses tersebut menghasilkan 8.086 records untuk diolah dengan *data mining*. Parameter *input* sejumlah lima buah, Sedangkan parameter *output*-nya adalah satu buah. Hal itu dapat dilihat pada tabel di berikut ini.

TABEL I. PARAMETER DATA INPUT DAN OUTPUT

No	Nama Parameter	Keterangan Nama Parameter	Uraian Paramater
1.	UMUR	Umur Wajib Pajak	
2.	STS_USAHA	Kode usaha status Wajib Pajak	1 : Tunggal 2 : Cabang 3 : Pusat Usaha 4 : Joint Operation 5 : Wanita Kawin 6 : Karyawan
3.	DAFTAR	Jangka Waktu Wajib Pajak Terdaftar	
4.	KD_WILAYAH	Kode wilayah	0:Kec. Palmerah 1: Kel. Palmerah 2 : Kel. Slipi 3:Kel. Kemanggisan

			4 : Kel. Jati Pulo 5:Kel. Bambu Utara 6:Kel. Bambu Selatan
5.	JENIS_KLU	Jenis Klasifikasi Lapangan Usaha	1 : Industri 2 : Jasa 3 : Perdagangan
6.	STS_LAPOR	Status Laporan SPT Tahunan PPh Orang Pribadi	1 : Laporan 2 : Tidak lapor

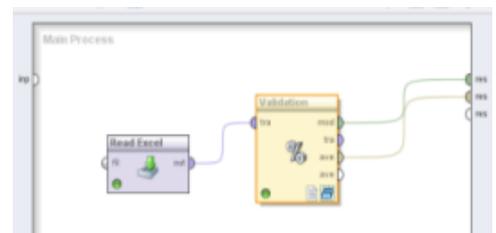
Dengan menggunakan metode pemilihan sampel *Stratified Sampling*, data tersebut digunakan sebagai *data training* dan *data testing*. Perbandingan yang digunakan untuk penentuan *data training* dan *data testing* adalah 90:10, yaitu 90% ditentukan sebagai *data training* dan 10% ditentukan sebagai *data testing*.

B. Proses Data Mining

Model *data mining* dalam penelitian ini dibuat dengan menggunakan perangkat lunak Rapid Miner 5.2. Algoritma berbasis jaringan syaraf tiruan (*artificial neural network*) yang berupa algoritma *Multilayer Perceptron* (MLP) dan algoritma *Support Vector Machine* (SVM) dijalankan dengan menggunakan perangkat lunak tersebut. Model yang dihasilkan dijelaskan sebagai berikut.

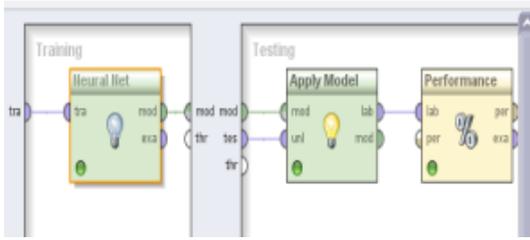
1) Algoritma *Multilayer Perceptron* (MLP)

Pembuatan model diawali dengan pembacaan file data (*Read Excell*). *Data training* dan *data testing* disimpan dalam satu file *Excell 2003*. Data tersebut kemudian divalidasi dengan perbandingan *data training* dan *data testing* adalah 90:10.



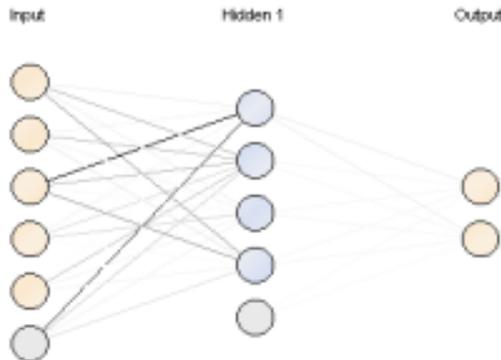
Gambar 1. Model Algoritma *Multilayer Perceptron* (MLP)-1

Langkah selanjutnya adalah penentuan model algoritmanya, dalam hal ini adalah *Naural Network*. Pada tahapan ini, diatur pembuatan *hidden layer* yang digunakan dalam algoritma.



Gambar 2. Model Algoritma *Multilayer Perceptron* (MLP)-2

Ketika model ini dijalankan, maka dapat dihasilkan arsitektur *Multilayer Perceptron* (MLP) sebagaimana dapat dilihat pada gambar di bawah ini.

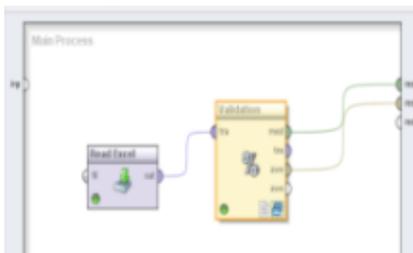


Gambar 3. Arsitektur *Multilayer Perceptron* (MLP)

Arsitektur *Multilayer Perceptron* (MLP) tersebut terdapat tiga lapisan, yaitu *input layer*, *hidden layer*, dan *output layer*. *Input layer* terdiri atas 6 *node* dengan 5 prediktor dan 1 label. *Hidden layer* terdiri atas 4 *node* dan 1 bias. Sedangkan *output layer* terdiri atas 2 *node*.

2) Algoritma *Support Vector Machine* (SVM)

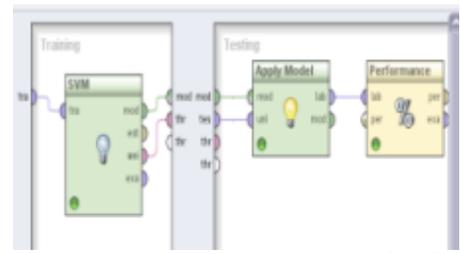
Seperti halnya pembuatan model algoritma *Multilayer Perceptron* (MLP), pembuatan model algoritma *Support Vector Machine* (SVM) diawali dengan pembacaan file data (*Read Excell*). *Data training* dan *data testing* disimpan dalam satu file *Excell 2003*. Data tersebut kemudian divalidasi dengan perbandingan *data training* dan *data testing* adalah 90:10.



Gambar 4. Model Algoritma *Support Vector Machine* (SVM)-1

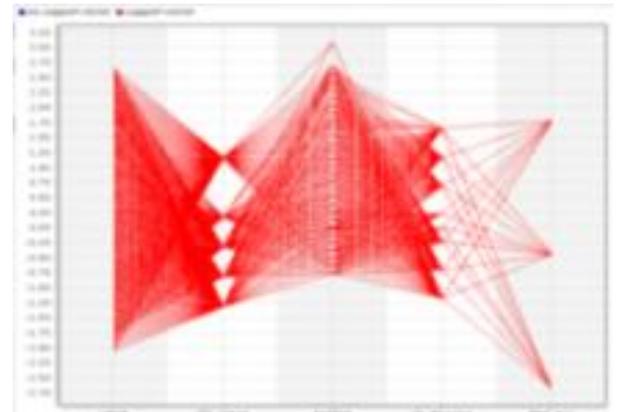
Langkah selanjutnya adalah penentuan model algoritmanya, dalam hal ini adalah *Support Vector*

Machine (SVM). Pada tahapan ini, diatur nilai *C* dan *epsilon* yang digunakan dalam algoritma.



Gambar 5. Model Algoritma *Support Vector Machine* (SVM)-2

Ketika model ini dijalankan, maka dapat dihasilkan arsitektur *Support Vector Machine* (SVM) sebagaimana dapat dilihat pada gambar di bawah ini.



Gambar 6. Arsitektur *Support Vector Machine* (SVM)

Pada gambar di atas terlihat kaitan antar variabel prediktor, yaitu *UMUR*, *STS_USAHA*, *DAFTAR*, *KD_WILAYAH*, dan *JENIS_KLU*. Semuanya berhubungan membentuk suatu *suport mechine* untuk menghasilkan data *output*.

IV. HASIL DAN DISKUSI

Kedua model algoritma yang dihasilkan oleh perangkat lunak Rapid Miner 5.2, dievaluasi dan diuji dengan *confusion matrix* dan kurva *Receiver Operating Characteristic* (ROC). Kedua model tersebut dapat dihasilkan dari Rapid Miner 5.2.

1) Model algoritma *Multilayer Perceptron* (MLP)

Confusion matrix dan kurva *Receiver Operating Characteristic* (ROC) yang dihasilkan dari model algoritma *Multilayer Perceptron* (MLP) dapat dijelaskan di bawah ini.

a) *Confusion matrix*

Data *Confusion matrix* untuk model algoritma *Multilayer Perceptron* (MLP) dapat dilihat pada tabel sebagai berikut.

TABEL II. *CONFUSION MATRIX* MODEL ALGORITMA *MULTILAYER PERCEPTRON (MLP)*

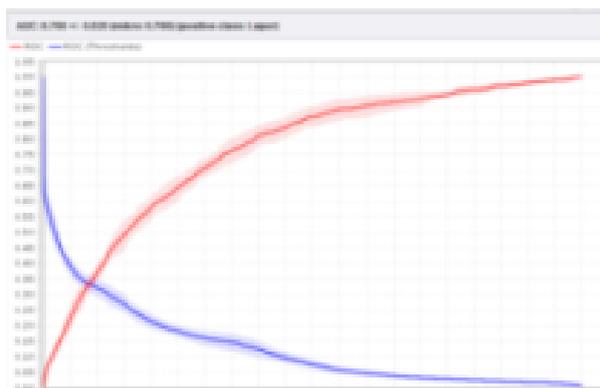
Accuracy 82.15%	True tdkLapor	True Lapor	Class precision
Pred.tdkLapor	6488	1308	83.22%
Pred Lapor	135	155	53.45%
Class recall	97.96%	10.59%	

Dari tabel tersebut terlihat bahwa jumlah data yang diolah adalah 8.086 *records* yang telah ditentukan perbandingan antara *data training* dan *data testing* adalah 90:10. Dapat diketahui bahwa 6.488 *records* diprediksi Tidak lapor, ternyata benar Tidak lapor. 135 *records* diprediksi Lapor, ternyata benar Tidak lapor. 1.308 *records* diprediksi Tidak lapor, ternyata benar Lapor. 155 *records* diprediksi Tidak lapor, ternyata benar Lapor.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{(6.488+155)}{(6.488+155+135+1.308)} = 0,8215$$

Sehingga data *confusion matrix* tersebut menghasilkan tingkat akurasi 82,15%.

- b) Kurva *Receiver Operating Characteristic (ROC)*
Selain *confusion matrix*, kurva *Receiver Operating Characteristic (ROC)* dihasilkan oleh Rapid Miner 5.2. Kurva tersebut dapat dilihat di bawah ini.



Gambar 7. Kurva *Receiver Operating Characteristic (ROC)* Model Algoritma *Multilayer Perceptron (MLP)*

Kurva *Receiver Operating Characteristic (ROC)* digunakan untuk mengekspresikan data *confusion matrix*. Garis horizontal mewakili nilai *false positives (FP)* dan garis vertikal mewakili nilai *true positives (TP)*. Dari gambar tersebut dapat diketahui bahwa nilai *The Area Under Curve (AUC)* model algoritma *Multilayer*

Perceptron (MLP) adalah 0,766. Hal ini menunjukkan bahwa dalam penelitian ini, model algoritma *Multilayer Perceptron (MLP)* mencapai klasifikasi cukup (*fair classification*).

- 2) Model algoritma *Support Vector Machine (SVM)*
Confusion matrix dan kurva *Receiver Operating Characteristic (ROC)* yang dihasilkan dari model algoritma *Support Vector Machine (SVM)* dapat dijelaskan di bawah ini.

- a) *Confusion matrix*
Data *Confusion matrix* untuk model algoritma *Support Vector Machine (SVM)* dapat dilihat pada tabel sebagai berikut.

TABEL III. *CONFUSION MATRIX* MODEL ALGORITMA *SUPPORT VECTOR MACHINE (SVM)*

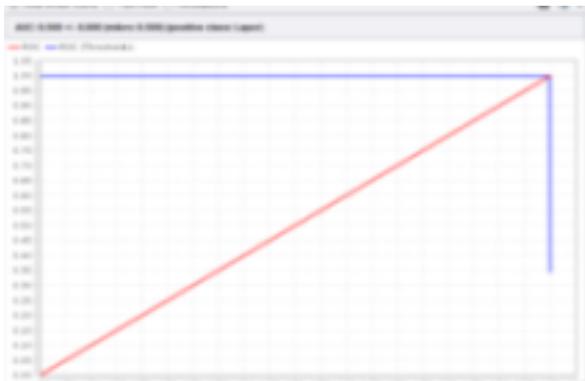
Accuracy 81.91%	True tdkLapor	True Lapor	Class precision
Pred.tdkLapor	6623	1463	81.91%
Pred. Lapor	0	0	0.00%
Class recall	100.00%	0.00%	

Dari tabel tersebut terlihat bahwa jumlah data yang diolah adalah 8.086 *records* yang telah ditentukan perbandingan antara *data training* dan *data testing* adalah 90:10. Dapat diketahui bahwa 6.623 *records* diprediksi Tidak lapor, ternyata benar Tidak lapor. 0 *records* diprediksi Lapor, ternyata benar Tidak lapor. 1.463 *records* diprediksi Tidak lapor, ternyata benar Lapor. 0 *records* diprediksi Tidak lapor, ternyata benar Lapor.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{(6.623+0)}{(6.623+0+1.463)} = 0,8191$$

Sehingga data *confusion matrix* tersebut menghasilkan tingkat akurasi 81,91%.

- b) Kurva *Receiver Operating Characteristic (ROC)*
Selain *confusion matrix*, kurva *Receiver Operating Characteristic (ROC)* dihasilkan oleh Rapid Miner 5.2. Kurva tersebut dapat dilihat pada Gambar 8.



Gambar 8. Kurva Receiver Operating Characteristic (ROC) Model Algoritma Support Vector Machine (SVM)

V. SIMPULAN

Dari penelitian dan pengujian yang sudah dilakukan dapat disimpulkan bahwa antara model algoritma *Multilayer Perceptron* (MLP) dan model algoritma *Support Vector Machine* (SVM), menghasilkan tingkat akurasi 82,15% dan nilai *The Area Under Curve* (AUC) 0,766. Sedangkan model algoritma *Support Vector Machine* (SVM) menghasilkan tingkat akurasi 81,91% dan nilai *The Area Under Curve* (AUC) 0,500. Sehingga terlihat bahwa model algoritma *Multilayer Perceptron* (MLP) menghasilkan tingkat akurasi lebih tinggi dibandingkan dengan model algoritma *Support Vector Machine* (SVM). Oleh karena itu, dalam penelitian ini, model algoritma *Multilayer Perceptron* (MLP) adalah model algoritma terpilih.

DAFTAR PUSTAKA

[1] Adams, H.C., "The Science of Finance", D. Appleton and Company, New York, 1898.
 [2] Arum, Harjanti Puspa, 2012, "Pengaruh Kesadaran Wajib Pajak, Pelayanan Fiskus, dan Sanksi Pajak terhadap Kepatuhan Wajib Pajak Orang Pribadi yang Melakukan Kegiatan Usaha dan Pekerjaan Bebas: Studi di Wilayah KPP Pratama Cilacap", Skripsi, Semarang: Universitas Diponegoro.
 [3] Bastable, C.F., "Public Finance", 3rd ed., Macmillan & Co., London, 1993.

[4] Brotodihardjo, R. Santoso, "Pengantar Ilmu Hukum Pajak", Eresco, Bandung, 1989.
 [5] Data Sekunder Kantor Pelayanan Pajak (KPP) Pratama Jakarta Palmerah.
 [6] Gaur, Priyanka, et.al, -, "Naural Network in Data Mining", International Journal of Electronics and Computer Science Engineering, ISSN-2277-1956, Vol. 1 No. 3: 1449-1453, Rajasthan.
 [7] Gorunescu, F., "Data Mining: Concepts, Models and Techniques", Springer-Verlag Berlin Heidelberg, Berlin, 2011.
 [8] Han, J., et.al, "Data Mining Concepts and Techniques", 2nd ed., Morgan Kauffman, San Fransisco, 2006.
 [9] Handayana, Prisma, 2012, "Penerapan Particle Swarm Optimization untuk Seleksi Atribut pada Metode Support Vector Machine untuk Prediksi Penyakit Diabetes", Tesis, Jakarta: Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri.
 [10] Larose, Daniel T., "Discovering Knowledge in Data: An Introduction to Data Mining", John Willey & Sons, Inc., New Jersey, 2005.
 [11] Myatt, Glen J., "Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining", John Wiley & Sons, Inc., New Jersey, 2007.
 [12] Nurmantu, Safri, "Pengantar Perpajakan", Granit, Jakarta, 2003.
 [13] Santoso, Budi, "Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori dan Aplikasi", cetakan pertama, Graha Ilmu, Yogyakarta, 2007.
 [14] Sardiarinto, 2012, "Komparasi Penerapan Algoritma C 4.5, K-Nearest Neighbor, dan Naive Bayes dalam Penentuan Kelayakan Peminjaman Kredit Nasabah Koperasi", Tesis, Jakarta: Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri.
 [15] Seligman, Edwin R.A., "Essays on Taxation", 10th ed., Macmillan Co., New York, 1925.
 [16] Setiadi, Ahmad, 2012, "Kajian Penerapan Model Neural Network untuk Prediksi Penyakit Hati", Tesis, Jakarta: Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri.
 [17] Soemitro, Rohmat, "Dasar-dasar Hukum Pajak dan Pajak Pendapatan 1944", cetakan ke-9, Eresco, Bandung, 1979.
 [18] Sunaryo, Dodik, 2007, "Penerapan Metode Circular Backpropagation pada Jaringan Syaraf Tiruan untuk klasifikasi pada Sistem Pengenalan Wajah", Tugas Akhir, Surabaya: Institut Teknologi Sepuluh November.
 [19] Witten, I. H., et.al, "Data Mining: Practical Machine Learning and Tools", 3rd ed., Morgan Kaufmann Publisher, Burlington, 2011.