KOMPARASI *SUPPORT VECTOR MACHINE* DAN *NEURAL NETWORK* UNTUK PREDIKSI KELULUSAN SERTIFIKASI BENIH KENTANG

Usep Tatang Suryadi
Komputer Akuntansi STMIK SUBANG
Jl. Marsinu No. 5 Subang
e-mail : ugie89@gmail.com

Abstrak
Sertifikasi benih merupakan masalah penting bagi petani karena merupakan salah satu upaya menjaga kualitas benih. Penelitian untuk memprediksi kelulusan sertifikasi benih dengan menggunakan teknik data mining masih jarang dilakukan, dari beberapa penelitian dengan konteks yang sama, yaitu penerapan metode data mining untuk klasifikasi kelulusan dengan jenis data dan kelas yang sama, yaitu dua kelas Lulus dan Tidak Lulus. Metode Neural Network memliliki performa yang baik dibandingkan dengan metode Support Vector Machine yang mampu memberikan solusi secara global optimum. Penelitian ini membandingkan akurasi metode Neural Network dan Support Vector Machine untuk menyelesaikan masalah prediksi kelulusan sertifikasi benih. Proses validasi menggunakan Split Validation, sedangkan pengujian model menggunakan metode Confusion Matrix dan ROC Curve. Hasil pengujian menunjukkan model dengan metode Neural Network memiliki akurasi sebesar 96.61% dan nilai AUC sebesar 0.997 sedangkan untuk metode Support Vector Machine memiliki nilai akurasi sebesar 98.91% dan nilai AUC sebesar 1.000. Sehingga dapat disimpulkan penerapan metode Support Vector Machine lebih baik dari Neural Network pada data sertifikasi benih kentang.

Kata Kunci: *Benih Kentang, Data Mining, Neural Network, Support Vector machine, Global Optimum, Split Validation, Confusion Matrix, AUC, ROC.*

1. PENDAHULUAN

1.1 Latar Belakang

1.1.1 Identifikasi Masalah General

Kentang merupakan salah satu alternatif makanan pokok yang dapat dibuat beraneka jenis panganan olahan dan food therapy bagi penderita diabetes, perawatan kecantikan, asam lambung dan lain sebagainya. Prediksi kebutuhan kentang dalam negeri berkisar 8,9 juta ton/tahun. Selama ini produksi kentang nasional masih $\pm$ 1,1 juta ton/tahun, termasuk kentang sayuran, dari luas panen 80.000 ha. Potensi ini masih perlu dikembangkan, karena potensi lahan yang berada pada ketinggian diatas 700 m dpl masih sangat luas. [1]

Benih merupakan salah satu kunci kesuksesan budidaya pertanian. Kebutuhan benih kentang nasional setiap tahunnya diprediksi mencapai 128. 613.000 ton dengan nilai Rp. 1,29 trilyun, jika harga benih Rp. 10.000/kg. Selama ini kebutuhan benih yang sehat dan bermutu baru dapat tercukupi sekitar 6.430 ton, termasuk import (Deptan, 2007). Harga benih import sangat mahal, dapat mencapai Rp. 20.000/kg untuk benih sebar (G4), sedang harga benih G4 untuk produksi dalam negeri (Balitsa, Pengalengan) mencapai 10.000/kg. Minim dan mahalnya benih yang tersedia menyebabkan petani enggan untuk menggunakan benih bermutu (bersertifikat) untuk dipakai sehingga produktivitas lahan kentang di Indonesia masih sangat rendah.[1]

Ketersediaan benih bermutu mayoritas komoditi di sektor hortikultura tidak sebanding dengan kebutuhan petani. Terbatasnya ketersediaan benih sumber komoditi hortikultura khususnya benih sayuran yang sesuai dengan kebutuhan pasar, merupakan kondisi dan permasalahan perbenihan horikultura yang dihadapi.[2]

BPSBTPH (Balai Pengawasan dan Sertifikasi Benih Tanaman Pangan dan Hortikultura) merupakan Unit Pelaksanaan Teknis Dinas Pertanian Tanaman Pangan Provinsi Jawa Barat yang mempunyai tugas pokok dan fungsi di bidang pelayanan pengawasan mutu dan sertifikasi kepada para

produsen penangkar/penyalur benih untuk menghasilkan benih bermutu meliputi: kegiatan kultivar, sertifikasi benih, pengujian mutu benih di laboratorium dan pengawasan mutu benih di pasaran. Kegiatan tersebut dalam rangka mendukung ketersediaan benih tanaman pangan dan hortikultura baik secara regional Jawa Barat maupun tingkat nasional. Data sertifikasi pada BPSB masih sebatas data historis saja belum banyak diolah dan dimanfaatkan untuk kemudian digali informasi lebih ataupun sebagai data latih pada metode klasifikasi.

1.1.2 Identifikasi Masalah Spesifik

Belum adanya penggalian informasi dari data sertifikasi yang ada pada BPSBTPH Jawa Barat, untuk mendapatkan manfaat lebih dari informasi yang dihasilkan.

Support Vector Machine, adalah suatu teknik yg relative baru untuk melakukan prediksi, baik untuk metode klasifikasi maupun regresi (1995).dalam hal fungsi dan kondisi permasalahan yang ada, SVM masih satu kelas dengan Neural Network, dimana keduanya merupakan supervised learning. Perbedaannya SVM menemukan solusi yang global optimal sedangkan neural network local optimal. Dalam SVM, kita berusaha mencari sebuah fungsi pemisah (*hyperplane*) yang optimal yang bisa memisahkan dua kelompok set data dari dua kelas yang berbeda. Dimana fungsi yang kita cari adalah fungsi linear yang bisa didefinisikan sebagai berikut:

$g(x) = sgn(f(x))$
Dengan $f(x) = w^T x + b$
Dimana x, w $\in$ R^n dan b $\in$ R.

Yang kita cari adalah set parameter (w,b), sehingga $f(x_i) = \langle x,w \rangle + b = y_i$ untuk semua i. dalam teknik SVM ini kita mencari *hyperplane* (fungsi pemisah/classifier) terbaik yang memisahkan dua macam objek/label/class. Mencari hyperplane terbaik ekuivalen dengan memaksimalkan margin atau jarak antara dua set obyek dari kelas yang berbeda. Jika $wx_1+b=+1$ adalah *hyperplane* pendukung dari kelas +1 dan $wx_2+b=-1$ adalah *hyperplane* pendukung dari kelas -1, maka margin dapat dihitung dengan mencari jarak kedua hyperplane pendukung. Sehingga

$(wx_1+b=+1)-(wx_2+b=-1) \Rightarrow w(x_1-x_2)=2 \Rightarrow \left(\frac{w}{||w||}(x_1 - x_2)\right) = \frac{2}{||w||}$

Untuk klasifikasi linier di dalam primal space, formulasi optimasi SVM adalah sebagai berikut:

$$min\frac{1}{2}||w||^2$$

dengan $y_i(wx_i + b) \geq 1$, $i = 1, .., l,$ dimana x_i adalah data input(variable parameter), y_i adalah data keluaran (variable kelas) dari x_i, w dan b adalah parameter-parameter yang kita cari nilainya.

Bila output data $y_i=+1$, maka fungsi pembatasnya adalah $(wx_i+b)\geq 1$,dan bila $y_i=-1$ maka $(wx_i+b)\leq -1$.
Dalam kasus yang tidak infeasible, dimana data tidak bis dikelompokan secara benar maka formulasinya adalah $min\frac{1}{2}||w||^2 + C\sum_{i=1}^{l} t_i$ dengan $y_i(wx_i+b)+t_i\geq 1$, $t_i\geq 0$, $i=1,\ldots,l$, dimana t_i adalah variable slack.
Berhubung data yang digunakan adalah data dengan dua kelas maka teknik yang digunakan adalah SVM linear.

Artificial Neural Network adalah sebuah teknik metode datamining yang meniru cara kerja dari jaringan syaraf (*neural network*) seperti yang telah disebutkan diatas, ANN juga merupakan metode yang *supervised learning* juga *unsupervised*. ANN ditentukan oleh tiga hal yaitu pola hubungan antar neuron, menentukan bobot penghubung dan fungsi aktivasi. [3]

Ada tiga lapis proses pada metode neural network, yang disebut *neural layers* yaitu lapisan input (menerima pola inputan data dari luar yang menggambarkan permasalahan), lapisan tersembuyni */hiden layer* dan lapisan output (merupakan solusi terhadap suatu permasalahan). Pada masalah yang penelitian ini, dilihat dari data yang ada, merupakan masalah yang sederhana dengan 2 kelas sehingga dipilih arsitektur ANN yang single layers network. Untuk formulasi ANN adalah sebagai berikut:
Jika net = $\sum x_i w_i$ maka fungsi aktivasinya adalah $f(net)=f(\sum x_i w_i)$.

Beberapa fungsi aktivasi yang digunakan adalah

- Fungsi threshold (batas ambang)

  f(x)= $\begin{cases} 1 \dots x \geq a \\ 0 \dots x \leq a \end{cases}$

  untuk kasus bilangan bipolar, 0 diganti dengan angka -1 sehingga persamaan diubah:

  f(x)= $\begin{cases} 1 \dots x \geq a \\ -1 \dots x \leq a \end{cases}$

- Fungsi sigmoid

  $$f(x) = {}^{1}/_{1+e^{-x}}$$

  Fungsi ini sering digunakan karena nilai fungsinya yang sangat mudah untuk didiferensiasikan, f(x)=f(x)(1-f(x))

- Fungsi identitas

  F(x)=x

  Digunakan jika keluaran yang dihasilkan oleh ANN merupakan sembarang bilangan real.

Confusion matrix [4] merupakan metode untuk mengevaluasi model klasifikasi pada data mining dengan menghasilkan nilai prediksi benar dan prediksi salah jika dibandingkan ke nilai tujuan (*target value*) dalam data.*Confusion matrix* adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan lima keluaran, yaitu: *recall, specificity, precision, negative predictive value dan acuraccy.*

Confusion Matrix		Target			
		Positive	Negative		
Model	Positive	a	b	*Positive Predictive Value*	a/(a+b)
	Negative	c	d	*Negative Predictive Value*	d/(c+d)
		Sensitivity	*Specificity*	**Accuracy** = (a+d)/(a+b+c+d)	
		a/(a+c)	d/(b+d)		

Gambar 1. Confusion Matrix

- ***Accuracy*** *:* proporsi jumlah prediksi yang benar
- ***Positive Predictive Value*** *or* ***Precision*** *:* proporsi kasus positif yang diidentifikasi dengan benar.
- ***Negative Predictive Value*** *:* proporsi kasus negatif yang diidentifikasi dengan benar.
- ***Sensitivity*** *or* ***Recall*** *:* proporsi kasus positif sebenarnya yang diidentifikasi dengan benar.
- ***Specificity*** *:* proporsi kasus negative sebenarnya yang diidentifikasi dengan benar.

ROC, kurva ROC dibagi menjadi dua dimensi, dimana tingkat True Positif di set pada sumbu *y* sedang tingkat False Positif pada sumbu *x*. Sedangkan untuk menentukan klasifikasi mana yang baik dipresentasikan dengan metode menghitung luas daerah di bawah kurva atau *AUC* yang diartikan sebagai probabilitas [5]. Luas daerah di bawah kurva mengukur kinerja diskriminatif dengan memperkirakan kemungkinan output dari sampel yang dipilih secara acakdari populasi negatif atau positif. Semakin besar nilai daerah bidang di bawah kurva maka semakin kuat klasifikasi yang digunakan. Dimana nilai AUC nilainya akan selalu antara 0,0 dan 1,0. Tabel berikut merupakan klasifikasi AUC.

Tabel Klasifikasi AUC

Performansi	Klasifikasi
0,99-1,00	Sangat Baik
0,80-0,90	Baik
0,70-0,80	Cukup
0,60-0,70	Rendah
0,50-0,60	Buruk

1.1.3 Analisis Masalah

Klasifikasi adalah suatu proses pengelompokan data yang didasarkan pada ciri-ciri tertentu kedalam kelas-kelas yang telah ditentukan. Klasifikasi juga merupakan proses pencarian sekumpulan model yang membedakan kelas data dengan tujuan agar dapat digunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya.[6]

1.1.4 Data set yang digunakan

Data yang digunakan adalah data sertifikasi benih kentang BPSBTPH JABAR (Balai Pengawasan dan Sertifikasi Benih Tanaman Pangan dan Hortikultura Jawa Barat) pada tahapan pemeriksaan lapangan. Dimana memiliki 7 atribut, 5 parameter dan 2 kelas seperti berikut:

Tabel 1. Tipe atribut

Virus	Polynominal	Atribut
Bakteri Layu	Polynominal	Atribut
Busuk Daun	Polynominal	Atribut
Aphid	Polynominal	Atribut
CVL	Polynominal	Atribut
Lulus	Binominal	Kelas
Tidak Lulus	Binominal	Kelas

Data ini merupakan data nonrentet waktu, dimana data tidak mengacu pada perguliran waktu. Dengan tipe atribut parameter seperti pada tabel di atas.

Data set yang digunakan

No	Pemeriksaan lapangan					
	Virus (%)	Bakteri Layu (%)	Busuk Daun (%)	Aphid (%)	CVL (%)	L/TL
1	0	0.7	0	0	0	TL
2	0	0.59	0	0	0	TL
3	0	0	0	0	0	L
4	0.24	0.06	0	0	0	L
5	0	0.7	0	0	0	TL
6	0	0.69	0	0	0	TL
7	0	0	0	0	0	L
8	0.24	0.06	0	0	0	L
9	0	0.7	0	0	0	TL
10	0	0.64	0	0	0	TL
11	0	0.07	0	0	0	L
12	0.12	0.12	0	0	0	L
13	0	0.7	0	0	0	TL
14	0	1.05	0	0	0	TL
15	0	0	0	0	0	L
16	0.12	0.6	0	0	0	L
17	0	0.7	0	0	0	TL
18	0	0.87	0	0	0	TL
19	0	0.7	0	0	0	TL
20	0	1.09	0	0	0	TL

2. TINJAUAN PUSTAKA

Beberapa penelitian terdahulu baik penelitian yang hanya menggunaan metode yang sama maupun yang juga pada bidang yang sama adalah sebagai berikut:

Penggunaan data mining dengan metode klasifikasi, Menggunakan data tahap pertumbuhan katun daerah pertanian katun/kapas India bagian selatan, dengan 24 atribut, tiga label kualitas yaitu baik, rata-rata dan buruk dengan jumlah 900 data sample, dimana tingkat akurasi Neural Network Multilayer Perceptron sebesar 98,78% menggunakan metode evaluasi 10-vold cross validation . [7]

Klasifikasi bibit katun/kapas juga dilakukan oleh, menggunakan data pertumbuhan katun/kapas pada daerah yang sama dengan penelitian sebelumnya, dengan jumlah data 500 kasus menggunakan 24 atribut dan tiga label untuk kualitas, untuk Neural Network Multi Layer perceptron menghasilkan nilai akurasi sebesar 98,78%. [8]

3. METODE PENELITIAN

Pada penelitian ini ada beberapa tahapan yang dilakukan sebagai berikut:

1. Studi literatur dan pengumpulan data set, di mana meliputi studi literature untuk referensi dalam penelitian berupa jurnal dan karya ilmiah yang relevan dengan penelitian. Data set yang digunakan merupakan data kelulusan sertifikasi benih pada tahap pemeriksaan lapangan dari Balai Pengawasan dan Sertifikasi Tanaman Panagn dan Hortikultura Jawa Barat (BPSBTPH JABAR), tahun 1996-1998 sejumlah 500 Data set.
2. Pengolahan data awal, meliputi data cleansing: membersihkan noise dan data yang tidak konsisten sehingga diperoleh 178 data training dan 90 data testing; data reduction: untuk menghapus atribut-atribut yang tidak diperlukan seperti nomor induk, nama pengaju, alamat, pemeriksaan media tanam dsb; normalisasi data: pada row yang atributnya tidak lengkap atau tidak terisi dilakukan handling missing value.
3. Pemodelan, setelah data benar-benar bersih dari *empty field*, *garbage,*dan *redundancy* baru bisa diolah untuk kemudian pemilihan tipe validasi dimana menggunakan validasi tipe *Split Validation* baru kemudian dilakukan ditentukan metode untuk pemodelan. Pemodelan pertama dilakukan dengan menggunakan metode *Support Vector Machine* baru kemudian membuat pemodelan lain dengan menggunakan metode *Neural Network.* Untuk selanjutnya dilakukan pengukuran *performansi* dari masing-masing metode menggunakan teknik *Confusion Matrix*.

4. HASIL DAN PEMBAHASAN

Setelah proses pengolahan data awal selesai, maka data set untuk penelitian siap untuk digunakan. Validasi menggunakan *Split Validation*. Pemilihan *Split validation* sendiri agar lebih memudahkan penelitian menentukan pembagian data *traning* dan *testing* dengan *ratio* yg bisa kita tentukan sendiri. Setelah pemilihan jenis validasi, kemudian memilih metode mana yang akan dipakai untuk pemodelan, dalam penelitian ini dipilih Neural Network dasar untuk penelitian pertama, kemudian pemilihan penerapan model dan pengujian performance. Begitupun untuk penelitian metode Support Vector Machine.

Improved Neural Net

Hidden 1
========
Node 1 (Sigmoid)

Virus (%): 1.664
Bakteri Layu (%): -5.336
Busuk Daun (%): -0.019
Aphid (%): -1.436
CVL(%): 0.227
Bias: -1.870

Node 2 (Sigmoid)

Virus (%): 2.274
Bakteri Layu (%): -6.860
Busuk Daun (%): -0.010
Aphid (%): -1.825
CVL(%): 0.334
Bias: -2.017

Node 3 (Sigmoid)

Virus (%): 3.030
Bakteri Layu (%): -8.787
Busuk Daun (%): 0.013
Aphid (%): -2.292
CVL(%): 0.413
Bias: -2.093

Node 4 (Sigmoid)

Virus (%): 1.284
Bakteri Layu (%): -4.519
Busuk Daun (%): -0.040
Aphid (%): -1.291
CVL(%): 0.169
Bias: -1.691

Node 5 (Sigmoid)

Virus (%): 1.752
Bakteri Layu (%): -5.530
Busuk Daun (%): -0.044
Aphid (%): -1.514
CVL(%): 0.281
Bias: -1.882

Output
======

Class 'TL' (Sigmoid)

Node 1: -2.666
Node 2: -3.548
Node 3: -4.784
Node 4: -2.178
Node 5: -2.796
Threshold: 3.699

Class 'L' (Sigmoid)

Node 1: 2.658
Node 2: 3.627
Node 3: 4.761
Node 4: 2.161
Node 5: 2.763
Threshold: -3.697

accuracy: 96.61%

	true TL	true L	class precision
pred. TL	7	2	77.78%
pred. L	0	50	100.00%
class recall	100.00%	96.15%	

Gambar 2. Performance metode Neural Network

PerformanceVector Machine
PerformanceVector:
accuracy: 98.31%
ConfusionMatrix:
True: TL L
TL: 7 1
L: 0 51
precision: 100.00% (positive class: L)
ConfusionMatrix:
True: TL L
TL: 7 1
L: 0 51
recall: 98.08% (positive class: L)
ConfusionMatrix:
True: TL L
TL: 7 1
L: 0 51
AUC (optimistic): 1.000 (positive class: L)
AUC: 1.000 (positive class: L)
AUC (pessimistic): 1.000 (positive class: L)

Tabel 2. Weight table

Virus (%)	0.2647069540890592
Bakteri Layu (%)	-1.5320620434216972
Busuk Daun (%)	0.0
Aphid (%)	-0.19816396011669252
CVL(%)	0.08543698916050788

	true TL	true L	class precision
pred. TL	7	1	87.50%
pred. L	0	51	100.00%
class recall	100.00%	98.08%	

Gambar 4. Tingkat akurasi SVM

5. KESIMPULAN

Dari hasil penelitian komparasi pemodelan data menggunakan di atas, setelah melakukan beberapa tahap pengolahan data awal, pemodelan metode dengan metode *Support Vector Machine* dan *Neural Network* kemudian pemodelan dan pengukuran preforma tiap tiap metode adalah sebagai berikut:

- *Support Vector Machine,* memiliki nilai *accuracy* sebesar 98,31% dan AUC sebesar 1.00
- *Neural Network,* memiliki nilai *accuracy* sebesar 96,61% dan AUC sebesar 0,997.

dengan metode *Support Vector Machine* memiliki akurasi yang lebih besar dibanding metode Neural Network. Sehingga dapat disimpulkan pada penelitian ini, SVM lebih baik didanding ANN pada data klasifikasi kelulusan sertifikasi benih.

DAFTAR PUSTAKA

[1] Baharuddin, Nurbaya, Kuswinanti, T., Lologau, B. A., (2011), *Effect of Clostridium spp in the Control of Ralstonia solanacearum on Potato Using Aerophonic Cultivated System*, alamat URL lengkap dapat diakses pada http://Frepository.unhas.ac.id/bitstream/handle/123456789/769/makalah/bahar.doc

[2] BPSBTPH P Jawa Barat Tahun 2012(2013). Laporan Tahunan kegiatan BPSBTPH Prov. Jawa Barat. Bandung.

[3] Pattiserlihun, A., dkk (2010). Aplikasi Jaringan Syaraf Tiruan (Artificial Neural Network) pada Pengenalan Pola Tulisan. Prosiding pertemeuan ilmiah XXV HFI Jateng & DIY.

[4] Sayad, S., Model Evaluation-Clasification. [online] 2013. (http://www.saedsayad.com/model_evaluation_c.htm).

[5] Florin Gorunescu (2011), Data Mining: Concepts, Model and Techniques, Prof. Janusz Kacprzyk and Prof. Lakhmi C. Jain, Eds. Berlin, Jerman: Springer. vol. 12.

[6] Mulyanto, A. (2009). "Sistem Informasi Konsep & Aplikasi". Yogyakarta: Pustaka Pelajar.

[7] Jamuna, K.S., Karvagavali, S., Vijaya, M.S. (2010). Classification of Seed Cotton Yield based on the Growth stages of Cotton crop using Machine Learning Techniques, IEEE International Journal.

[8] Revathi, P., Hemalatha, M., (2011). Categorize the Quality of Cotton Seeds Based on the Different Germination of the Cotton Using Machine Knowledge Approach, International Journal of Advanced Science and Technology Vol. 36, November, 2011.

[9] Hastuti, K., (2012). Analisis Komparasi Algoritma Klasifikasi Data Mining untuk PrediksivMahasiswa Non Aktif, Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012 (Semantik 2012). Semarang, 23 Juni 2012.