

ALGORITMA SUPPORT VECTOR MACHINE

by Purwanto Purwanto

Submission date: 18-Apr-2020 07:36PM (UTC+0700)

Submission ID: 1300907059

File name: Nas_Jurnal_101_Algoritma_Support_Vector_Machine.pdf (4.95M)

Word count: 2053

Character count: 12707

ALGORITMA SUPPORT VECTOR MACHINE UNTUK MEMPREDIKSI NILAI UJIAN NASIONAL

Emi Rizky¹, Purwanto², H. Himawan³

¹²³Pascasarjana Teknik Informatika Universitas Dian Nuswantoro

ABSTRACT

In order to improve the quality of graduate education through an exam in order to compete in domestic, regional and international levels and therefore require the achievement of national standards through the National Examination (UN). produce test scores that boast with the title and can pass the National Exam, due to lack of graduates when the National Examination become routine issues annually. This problem is felt by students, parents, teachers, educational units and agencies associated with other national exams. By looking at the reason we need a prediction to predict the value of the UN. Soft computing has several abilities one of which is a technique that can be used to predict the ability of students to acquire the methods of the National Examination Support Vector Machine (SVM) which is a branch of artificial intelligence where the processing system configuration information obtained performance model for the prediction of the National Examination the Root mean squared Error (RMSE) is the best for Indonesian was $0.713 + / - 0.173$, English at $0586 + / - 0.066$, and Mathematics by $0882 + / - 0188$. configuration with predictions using a barometer. k -fold 10, C (cost) of 0.1 with kernel-type radial Indonesian subjects, k -fold 10, C (cost) of 0.3 with radial kernel type for the subjects of English and Mathematics.

Keywords: National Examination, SVM (Support Vector Machine), RMSE (root mean error squared)

13

1. PENDAHULUAN

1.1. Latar Belakang

Dalam rangka pengendalian mutu pendidikan secara nasional dilakukan evaluasi sebagai bentuk akuntabilitas penyelenggara pendidikan kepada pihak-pihak yang berkepentingan oleh lembaga yang mandiri secara berkala, menyeluruh, transparan dan sistemik, untuk menilai pencapaian standar nasional pendidikan dan proses pemantuan evaluasi. Ujian sebagai kegiatan pengukuran dan penilaian pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dalam kelompok mata pelajaran ilmu pengetahuan dan teknologi. Mata pelajaran yang diujikan dalam Ujian Nasional di Sekolah Menengah Kejuruan adalah mata pelajaran yang berperan penting dalam kelulusan yaitu Bahasa Indonesia, Bahasa Inggris dan Matematika. Melihat betapa pentingnya UN sebagai penentu kelulusan, oleh sebab itu dalam sistem belajar mengajar perlu adanya antisipasi sebelumnya agar bisa mengetahui siswa yang kurang atau tidak lulus dalam pelajaran Ujian Nasional tersebut untuk kita drill diberi training mantapan untuk bisa penguasaan materi mata pelajaran yang kurang tersebut. Karena ketidaklulusan ketika mengikuti Ujian Nasional menjadi permasalahan rutin tiap tahun. Permasalahan ini dirasakan oleh siswa, orang tua siswa, guru, satuan pendidikan dan instansi terkait dengan ujian Nasional lainnya.

Beberapa penelitian terkait telah dikaji diantaranya penelitian Chun-Tek Lye dkk [1]. Penelitian ini menggunakan metode Back-Propagansi Neural Network (BPNN), Regresi Pohon (CART), dan Generalized Regression Neural Network (GRNN) untuk memprediksi prestasi matematika siswa pra-universitas. Dengan menggunakan data pendaftaran, penelitian ini bertujuan untuk memprediksi hasil evaluasi tengah semester siswa pada tahap awal studi mereka. Hasil evaluasi dengan tingkat tertinggi dari 66,67% dan RMSE terendah dari 0,7174, serta dalam memprediksi hasil ujian akhir dengan hit rate tertinggi 71,11% dan RMSE terendah 0,4733. Selanjutnya, akurasi dan RMSE dari semua model juga menunjukkan perbaikan dari kisaran 0,84% menjadi 6,66%, dan 0,2173-0,2574 masing-masing ketika

variabel baru-baru ini termasuk dalam model memprediksi. Guna memperbaiki tingkat kesalahan (*error*) penelitian sebelumnya, maka penulis mengajukan model prediksi kelulusan dengan menggunakan metode SVM.

Penulis mempunyai alasan karena SVM (*Support Vector Machine*) memiliki landasan teori yang dapat dianalisis dengan jelas, dan SVM (*Support Vector Machine*) dapat diimplementasikan relative mudah, karena proses penentuan SVM dapat dirumuskan dalam standar program kuadrat (*quadratik programming / QP problem*).

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas maka permasalahan utama dalam penelitian adanya adalah adanya kesulitan dalam memprediksi kemampuan dan kematangan siswa dalam menghadapi Ujian Nasional agar dapat direncanakan bimbingan yang lebih intensif dan maksimal sehingga dapat dicapai nilai ujian dengan predikat yang membanggakan. Secara spesifik masalah penelitian ini (*research gap*) adalah masih tingginya *error* yang dihasilkan oleh model-model prediksi sebelumnya.

1.3. Tujuan

Penelitian ini bertujuan untuk menemukan model dengan performa terbaik prediksi nilai Ujian Nasional (UN) bagi siswa SMK tahun pelajaran yang akan datang berdasarkan nilai mata pelajaran Bahasa Indonesia, Bahasa Inggris, dan Matematika.

1.4. Manfaat

Hasil penelitian ini dapat memberikan informasi lebih awal tentang prediksi nilai Ujian Nasional siswa SMK PGRI Kota Tegal, dalam mengikuti UN tahun yang akan datang, sehingga pihak sekolah dapat mengambil langkah-langkah dalam mengatur strategi proses pembelajaran dan memberikan pemantapan bimbingan yang lebih intens kepada siswa, terutama siswa yang diidentifikasi mempunyai nilai rendah dalam ujian nasional pada mata pelajaran yang diujikan yaitu Bahasa Indonesia, Bahasa Inggris dan Matematika. Memberikan sumbangan bagi pengembangan teori yang berkaitan dengan prediksi nilai prestasi siswa dengan menggunakan SVM (*Support Vector Machine*) sebagai referensi bagi peneliti selanjutnya, khususnya penelitian yang berkaitan dengan prediksi dengan mengembangkan implementasi metode algoritma SVM (*Support Vector Machine*). Sebagai kebijakan penelitian yang diharapkan dapat menjadi sumbangan pengetahuan dan pengembangan dalam mengimplementasikan algoritma SVM (*Support Vector Machine*) untuk memprediksi berbagai masalah, sehingga nantinya dapat memberikan solusi terbaik berdasarkan analisis yang dapat dipertanggungjawabkan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terkait

Nuraeni [2] melakukan penelitian menggunakan jaringan syaraf tiruan untuk mengukur tingkat korelasi antara NEM dengan IPK kelulusan siswa. Selama ini NEM digunakan untuk seleksi penerimaan siswa baru, khususnya yang masuk melalui jalur *fellowship*, karena nilai NEM yang tinggi ternyata tidak menjamin lulusan berprestasi akademik tinggi, maka penelitian ini akan diuji korelasinya antara NEM SMA dengan indeks prestasi kumulatif lulusan. Hasil penelitian ini menunjukkan bahwa jaringan syaraf tiruan *backpropagation* memiliki tingkat keberhasilan 64% dalam memetakan komposisi NEM dan prestasi akademik lulusan.

Hilda Amalia [3] Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Ketepatan Kelulusan Mahasiswa Algoritma C4.5, Naïve Bayes dan Neural Network. Hasilnya C4.5 (33%), Naive Bayes (69.72%), Neural Network (78.83%). Triswanto [4] Komparasi Pemodelan Data Menggunakan *Support Vector Machine* dan Neural Network Untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa. Hasil pengujian menunjukkan *Support Vector Machine* memiliki akurasi sebesar 81.17% dan nilai 0.882 sedangkan Neural Network memiliki akurasi sebesar 78.83% dan nilai sebesar 0.858.

Redho Pati [5], Kekurangan dari metode *Support Vector Machine* ini adalah sebagai berikut : Sulit dipakai dalam problem berskala besar. Skala besar dalam hal ini dimaksudkan dengan jumlah sample yang diolah. SVM secara teoritik dikembangkan untuk problem klasifikasi dengan dua *class*. Sufandi [6] menggunakan Jaringan Syaraf Tiruan Multi layer Perceptron sebagai pengklasifikasian untuk memprediksi pengembangan sistem kemajuan mahasiswa pada Universitas Terbuka dengan sistem pembelajaran jarak jauh generalisasi terbaik hasil penelitian ini adalah dengan menggunakan data *training* 50%, data validasi 25% dan data *testing* 25% dengan jumlah node masukan 17, node tersembunyi 5 node keluaran 2 pada laju pembelajaran 0,01 yaitu dengan tingkat generalisasi tertinggi terhadap data validasi sebesar 97,27% dan protipe sistem kategori ‘aktif’ yaitu 95,88% sampai dengan 96,63%.

Ernawati [7] penelitian ini menggunakan algoritma C5.0 dan K-Nearest Neighbour (KNN) untuk prediksi status keaktifan studi mahasiswa perguruan tinggi swasta di Jakarta dengan variabel IPK yang menentukan potensi Mahasiswa aktif atau tidak pada waktu yang akan datang. Dalam percobaan dihasilkan Mahasiswa yang tidak aktif adalah Mahasiswa dengan IPK di bawah 1,77. Rata-rata keberhasilan algoritma C5.0 dan KNN dalam melakukan klasifikasi mencapai akurasi di atas 90% ini menunjukkan keduanya memiliki performa yang handal dalam melakukan klasifikasi.

Chun-Tek Lye dkk [1] penelitian ini menggunakan metode Back-Propagansi Neural Network (BPNN), Regresi Pohon (CART), dan Generalized Regression Neural Network (GRNN) untuk memprediksi prestasi matematika siswa pra-universitas. Dengan menggunakan data pendaftaran, penelitian ini bertujuan untuk memprediksi hasil evaluasi tengah semester siswa pada tahap awal studi mereka. Hasil evaluasi dengan tingkat tertinggi dari 66,67% dan RMSE terendah dari 0,7174, serta dalam memprediksi hasil ujian akhir dengan hit rate tertinggi 71,11% dan RMSE terendah 0,4733. Selanjutnya, akurasi dan RMSE dari semua model juga menunjukkan perbaikan dari kisaran 0,84% menjadi 6,66%, dan 0,2173-0,2574 masing-masing ketika variabel baru-baru ini termasuk dalam model memprediksi.

Baha Sen & Emine Ucar [8] penelitian ini menggunakan data mining Artificial Neural Network dan Decision Trees untuk mengevaluasi prestasi siswa dengan sistem pendidikan jarak jauh menurut kriteria seperti usia, jenis kelamin, jenis kelulusan SMA dan apakah mahasiswa yang pendidikan dengan belajar jarak jauh atau pendidikan reguler. Di antara jenis dua model, algoritma pohon keputusan menghasilkan hasil prediksi terbaik dengan Akurasi keseluruhan 97,8107% pada 10 kali lipat dataset ketidaksepakatan. Model pohon keputusan diikuti oleh jaringan saraf tiruan dengan akurasi keseluruhan 94,3752%.

Melihat hasil beberapa peneliti sebelumnya, disimpulkan prediksi dengan menggunakan metode *Support Vector Machine* (SVM) mempunyai daya prediksi dengan ketepatan yang tinggi, memiliki karakteristik dan kemampuan pembelajaran yang baik, maka penulis menggunakan metode *Support Vector Machine* sebagai uji coba dalam penelitian prediksi nilai Ujian Nasional.

2.2. Landasan Teori

2.2.1 Data Mining

Hal penting yang terkait dengan *data mining* adalah:

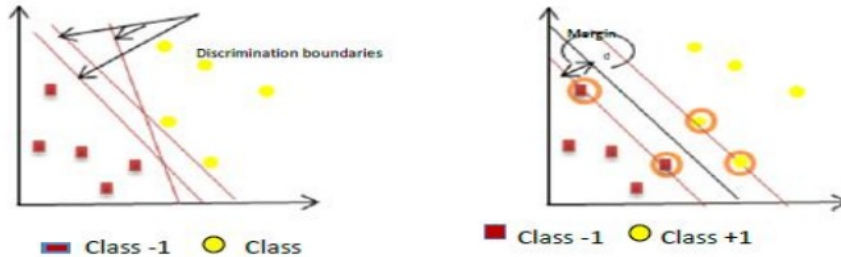
- a. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
- b. Data yang akan diproses berupa data yang sangat besar

Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat [9].

2.2.2 Algoritma *Support Vector Machine* (SVM)

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah *class* pada *input space*. Untuk *n*-dimensional space, *input data* x_i ($i=1, \dots, k$), dimana milik kelas 1 atau kelas 2 dan label yang terkait menjadi -1 untuk kelas 1 dan +1 untuk kelas 2. Gambar 2.1 memperlihatkan beberapa pattern yang merupakan anggota dari dua buah *class*: positif (dinotasikan dengan +1) dan negatif (dinotasikan dengan -1). Pattern yang tergabung pada *class* negatif disimbolkan dengan kotak, sedangkan pattern pada *class* positif, disimbolkan dengan lingkaran. Jika data *input* dapat dipisahkan secara linear, pemisahan *hyper plane* dapat diberikan dalam: Proses pembelajaran

dalam problem klasifikasi diterjemahkan sebagai upaya menemukan garis (*hyperplane*) yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (*discrimination boundaries*) ditunjukkan pada Gambar 1.



Gambar 1. SVM Berusaha Menemukan *Hyperplane* Terbaik yang Memisahkan Kedua *Class* Negatif dan Positif 2 [19]

Hyperplane pemisah terbaik antara kedua *class* dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperplane* tersebut dengan data terdekat dari masing-masing *class*. *Subset data training set* yang paling dekat ini disebut sebagai *support vector*. Garis solid pada Gambar 2.1 menunjukkan *hyperplane* yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua *class*, sedangkan titik kotak dan lingkaran yang berada dalam lingkaran hitam adalah *support vector*. Upaya mencari lokasi *hyperplane* optimal ini merupakan inti dari proses pembelajaran pada SVM [10].

Data yang tersedia dinotasikan sebagai $x \in R^d$ sedangkan label masing-masing dinotasikan $y_i \in \{-1, +1\}$ untuk $i = 1, 2, \dots, l$ yang mana l adalah banyaknya data. Diasumsikan kedua *class* -1 dan $+1$ dapat terpisah secara sempurna oleh *hyperplane* berdimensi d , yang didefinisikan:

Diasumsikan kedua *class* -1 dan $+1$ dapat terpisah secara sempurna oleh *hyperplane* berdimensi d , yang didefinisikan:

$$\vec{w} \cdot x + b = 0 \dots\dots\dots (2.1)$$

Sebuah *pattern* x_i yang termasuk *class* -1 (sampel negatif) dapat dirumuskan sebagai *pattern* yang memenuhi pertidaksamaan:

$$\vec{w} \cdot x + b = -1 \dots\dots\dots (2.2)$$

sedangkan *pattern* yang termasuk *class* $+1$ (sampel positif):

$$\vec{w} \cdot \vec{x} + b = +1 \dots\dots\dots (2.3)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu $1/||w||$. Hal ini dapat dirumuskan sebagai *Quadratic Programming* (QP) problem, yaitu mencari titik minimal dengan memperhatikan *constraint* :

$$\min_w \tau(w) = \frac{1}{2} - ||w||^2 \dots\dots\dots (2.4)$$

$$y_1 - (x_1 \cdot w + b) - 1 \geq 0, \forall_1 \dots\dots\dots (2.5)$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, diantaranya *Lagrange Multiplier* sebagaimana ditunjukkan pada persamaan di bawah ini :

$$L(w, b, \alpha) = \frac{1}{2} - \frac{1}{2} \|w\|^2 - w \sum_{i=1}^l \alpha_i (y_i((x_i \cdot w + b) - 1)) \quad (i=1,2,\dots,l) \dots\dots\dots (2.6)$$

α_i adalah *Lagrange multipliers*, yang bernilai nol atau positif ($\alpha_i \geq 0$). Nilai optimal dari persamaan diatas dapat dihitung dengan meminimalkan L terhadap w dan b, dan memaksimalkan L terhadap α_i . Dengan memperhatikan sifat bahwa pada titik optimal gradient L=0, persamaan dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung α_i saja, sebagaimana persamaan :
Maximize :

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i \cdot y_j \cdot \vec{x}_i \cdot \vec{x}_j \dots\dots\dots (2.7)$$

$$\alpha_i \geq 0 \quad (i = 1, 2, \dots, l) \quad \sum_{i=1}^l \alpha_i y_i = 0 \dots\dots\dots (2.8)$$

Dari hasil dari perhitungan ini diperoleh α_i yang kebanyakan bernilai positif. Data yang berkorelasi dengan α_i yang positif inilah yang disebut sebagai *support vector*.

2.2.3 RMSE (Root mean square Error)

Keakuratan keseluruhan dari setiap model peramalan baik itu rata-rata bergerak, *exponential smoothing* atau lainnya dapat dijelaskan dengan membandingkan nilai yang diproyeksikan dengan nilai aktual atau nilai yang diamati. Untuk tingkat akurasi peramalan dapat diukur dari nilai berikut, adapun rumus dari RMSE adalah sebagai berikut :

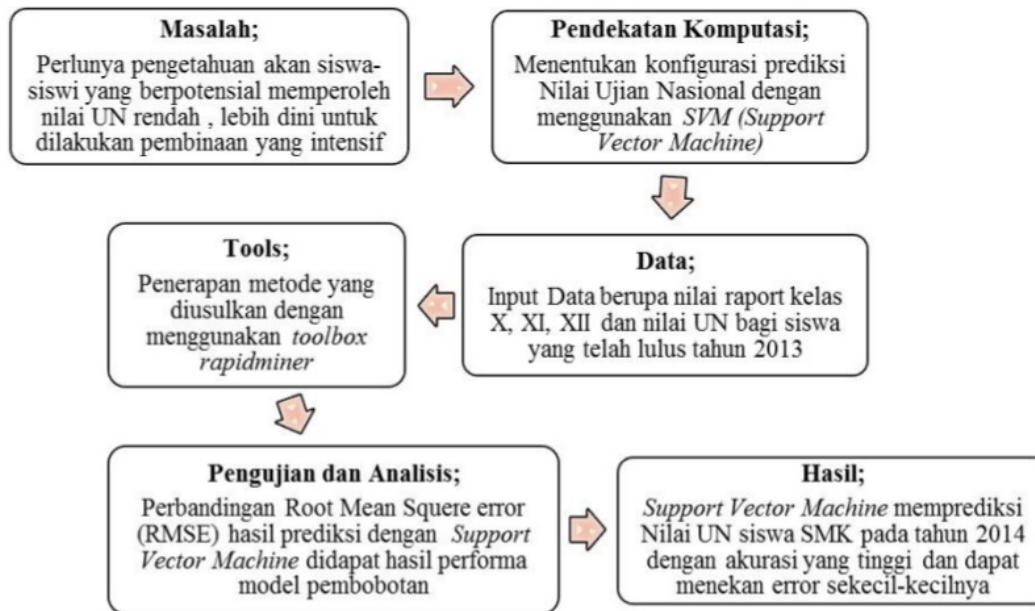
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_t - \hat{Y}_t)^2} \dots\dots\dots (2.9)$$

- N = Jumlah data
- Y_t = Data Riil
- \hat{Y}_t = Data Prediksi

Semakin kecil nilai RMSE, semakin baik tingkat akurasi prediksinya. Keakuratan sebuah model peramalan dalam melakukan prediksi ditentukan oleh nilai terkecil dari masing-masing metode akurasi data, semakin kecil nilai tersebut semakin akurat sebuah model melakukan prediksi. Untuk mengetahui teknik mana yang terbaik untuk data tertentu maka biasanya dilakukan dengan mencoba beberapa teknik berbeda dan memilih salah satu yang terbaik.

3. METODE PENELITIAN

Pada penelitian ini data yang digunakan adalah data nilai siswa siswi SMK PGRI Kota Tegal. Data nilai tersebut akan diolah menggunakan metode *data mining* untuk memprediksi nilai Ujian nasional. Tahapan penelitiannya sebagai berikut.



Gambar 2. Diagram Alir Metode Penelitian

3.1. Pengumpulan Data

Data awal yang digunakan untuk melakukan perhitungan dan analisis adalah dimulai dengan melakukan pengumpulan data. Bentuk data berupa data nilai siswa, yang diperoleh dari Wakil Kepala Sekolah bagian kurikulum SMK PGRI Kota Tegal, yaitu berupa data Nilai raport kelas X, XI dan XII dari semester 1 sampai 5 dan nilai Ujian Nasional Tahun ajaran 2012/2013 yang digunakan sebagai dasar dalam memprediksi nilai calon peserta ujian Nasional (UN) tahun 2013, Metode pengumpulan data dalam penelitian ini menggunakan Variabel-variabel sebagai prediksi nilai yaitu :

- Data siswa khususnya kelas XII Tahun Ajaran 2012 s/d 2013
- Nilai Raport semester 1- 5 mulai tahun ajaran 2010 s/d 2013 untuk matapelajaran yang di UN-kan yaitu Bahasa Indonesia, Bahasa Inggris, Matematika.
- Nilai Ujian Nasional 2013 untuk tiap matapelajaran yang di UN-kan juga.

3.2. Pengolahan Data Awal

Pengolahan data awal (*pre-processing*) dilakukan dengan pemilihan data dengan mencermati data sudah dalam bentuk mengolah data dengan menggunakan *Support Vector Machine* (SVM), data terlebih dahulu diolah dalam bentuk Microsoft Excel, kemudian data dilakukan pengecekan apakah ada data yang tidak valid atau masih salah atau masih ada data yang masih dalam bentuk teks, karena jika data masih dalam bentuk teks akan muncul kosong atau kesalahan (*missing value*) dan *outlier*.

Berdasarkan data tersebut dengan variabel yang sudah ada diolah dengan menggunakan *tool RapidMiner* dengan metode yang digunakan adalah hanya memakai *Support Vector Machines*, Penggunaan metode *Support Vector Machines* akan menghasilkan nilai akurasi yang lebih tinggi daripada metode yang lain.

3.3. Eksperimen

Penelitian ini menggunakan variabel yang terdiri dari 5 variabel *input* dan 1 variabel target yaitu sebagai *input* terdiri dari X1 untuk nilai raport semester 1, X2 untuk nilai raport semester 2, X3 untuk nilai raport semester 3, X4 untuk nilai raport semester 4, X5 untuk nilai raport semester 5 dan sebagai target nilai Y untuk nilai Ujian Nasional (UN).

Untuk skenario eksperimen ini uji coba dilakukan dengan melatih data dengan menggunakan iterasi/epoch yang berbeda-beda yaitu data yang digunakan menggunakan nilai raport semester 1 sampai dengan semester 5 dan nilai Ujian nasional siswa yang sudah lulus pada tahun pelajaran 2012/2013 sebagai data *training* sedangkan data yang diprediksi adalah nilai UN siswa kelas XII tahun pelajaran 2013/2014 sebagai target penelitian atau pembelajaran dalam menentukan model atau pola *SVM (Support Vector Machine)* dimana data prediksi yang diambil dari nilai semester 1 sampai dengan 5.

Setiap pasangan data *input output* akan dihitung tingkat *error* program yang diperoleh dari selisih antara output yang ideal dengan *output* yang dihasilkan oleh jaringan dan kemudian pada setiap akhir iterasi/epoch dihitung RMSE (*root mean square error*) semakin kecil dari nilai rata-rata *error* dan rata-rata RMSE menunjukkan bahwa aplikasi mampu berjalan sesuai dengan yang diharapkan dalam eksperimen.

3.4. Tahap Pengujian

Tahap pengujian dilakukan untuk mengetahui apakah *SVM (Support Vector Machine)* dapat mengenali pola dengan memberikan *input* yang mungkin berbeda dengan *input* pada tahap coba. Dalam penelitian uji coba ini digunakan *tools* rapidminer yang digunakan untuk menyelesaikan analisis dengan metode yang dipakai yaitu *SVM (Support Vector Machine)* dimana data yang dipakai adalah data *training* dan data *testing*. Kedua data ini sama-sama diinputkan ke dalam *input* model *SVM (Support Vector Machine)* dan pembagian data diolah oleh sistem secara otomatis hanya dengan mengatur *k-fold validation* pada default rapidminer.

Inputan data sebagai data pembentukan pola pembobotan nilai yang dapat menghasilkan metode *SVM (Support Vector Machine)* yang terbaik, sehingga dapat menghasilkan RMSE terkecil dengan waktu komputasi terpendek.

Medel performa ini nantinya digunakan untuk pengujian prediksi nilai siswa pada Ujian Nasional tahun 2014. Dengan memberikan masukan nilai semester 1 sampai dengan semester 5 untuk mata pelajaran Bahasa Indonesia, Bahasa Inggris, dan Matematika yang di UN-kan.

Tahapan eksperimen dengan membuat beberapa model konfigurasi pada rapidminer dengan mengatur default-nya yaitu *k-fold validation* merupakan *stratified 10-fold cross-validation* artinya bahwa *k-fold* ini adalah metode evaluasi standart, Nilai cost (C) dipilih untuk mengontrol keseimbangan antara nilai margin dan *error* prediksi. Semakin besar nilai C, maka nilai tetap yang diberikan terhadap data *error* juga semakin besar dan *type kernel* mengatasi kasus ketidaklinearan atau data yang sulit dipisahkan secara linier, fungsi *kernel* ini akan menentukan *feature space* fungsi prediksi yang akan dicari.

4. HASIL PENELITIAN DAN PEMBAHASAN

Algoritma *SVM* dengan *tool rapidminer* nilai parameter default sudah tersetting *type kernel* dengan pilihan *dot*, *radial* dan *polynomial*, nilai C (*cost*) dengan pilihan berdasarkan 0,1, 0,2, dan 0,3 dengan *k-fold (number of validation)* range $k=3$, $k=5$ dan $k=10$, akan tetapi nilai tersebut dapat dirubah sesuai dengan keinginan peneliti. Perubahan pengaturan bertujuan untuk mengatur *tool rapidminer* agar dapat diproses sesuai dengan komputasi yang peneliti kehendaki, sebagai contoh adalah data dibagi (10) bagian untuk data *testing* dan sebagian (90) sebagian digunakan untuk *learning*.

4.1. Prediksi Nilai UN Mapel Bahasa Indonesia

Tabel 1. Hasil Percobaan SVM Nilai UN Mapel Bahasa Indonesia

C (Cost)	k-fold								
	3			5			10		
KERNEL	0,1	0,2	0,3	0,1	0,2	0,3	0,1	0,2	0,3
Dot	1.152 +/- 0.553	1.170 +/- 0.576	1.208 +/- 0.627	1.050 +/- 0.687	1.048 +/- 0.687	1.047 +/- 0.689	0.919 +/- 0.731	0.924 +/- 0.738	0.925 +/- 0.742
Radial	0.744 +/- 0.071	0.738 +/- 0.069	0.741 +/- 0.066	0.733 +/- 0.112	0.732 +/- 0.107	0.737 +/- 0.105	0.713 +/- 0.173	0.720 +/- 0.167	0.730 +/- 0.164
Polynomial	1.313 +/- 0.580	1.302 +/- 0.576	1.312 +/- 0.582	0.853 +/- 0.105	0.851 +/- 0.096	0.877 +/- 0.088	0.787 +/- 0.187	0.800 +/- 0.197	0.806 +/- 0.198
k-fold	RMSE			C			Kernel		
3	0.738 +/- 0.069			0,2			Radial		
5	0.732 +/- 0.107			0,2			Radial		
10	0.713 +/- 0.173			0,2			Radial		

Dari eksperimen didapat RMSE paling kecil untuk mata pelajaran Bahasa Indonesia didapat dari kernel type radial dan nilai C(cost) 0,3 yaitu 0.713 +/- 0.173

Desain Prediksi mapel Bahasa Indonesia

- Total number of Support Vectors: 207
 - Bias (offset): 7.377
 - $w[X1] = 2.656$
 - $w[x2] = 1.005$
 - $w[x3] = 2.862$
 - $w[x4] = 0.089$
 - $w[x5] = -0.623$
- Artinya : $w[x1]$ = bobot nilai pengaruh dari $w[x1]$ terhadap Y sebesar 2.656

4.2. Prediksi Nilai UN Mapel Bahasa Inggris

Tabel 2 Hasil Percobaan SVM Nilai UN Mapel Bahasa Inggris

C (Cost)	k-fold								
	3			5			10		
KERNEL	0,1	0,2	0,3	0,1	0,2	0,3	0,1	0,2	0,3
Dot	0.724 +/- 0.192	0.718 +/- 0.186	0.721 +/- 0.187	0.754 +/- 0.353	0.770 +/- 0.384	0.786 +/- 0.421	0.707 +/- 0.408	0.710 +/- 0.436	0.718 +/- 0.456
Radial	0.614 +/- 0.034	0.608 +/- 0.026	0.605 +/- 0.013	0.600 +/- 0.050	0.591 +/- 0.049	0.589 +/- 0.045	0.591 +/- 0.079	0.586 +/- 0.070	0.586 +/- 0.066
Polynomial	0.656 +/- 0.012	0.664 +/- 0.016	0.671 +/- 0.022	0.655 +/- 0.066	0.668 +/- 0.068	0.676 +/- 0.070	0.815 +/- 0.573	0.648 +/- 0.114	0.773 +/- 0.423

<i>k-fold</i>	<i>RMSE</i>	<i>C</i>	<i>Kernel</i>
3	0.605 +/- 0.013	0,3	<i>Radial</i>
5	0.589 +/- 0.045	0,3	<i>Radial</i>
10	0.586 +/- 0.066	0,3	<i>Radial</i>

Dari eksperimen didapat RMSE paling kecil untuk mata pelajaran Bahasa Inggris didapat dari *kernel type radial* dan nilai C(cost) 0,3 yaitu 0.586 +/- 0.066

Desain Prediksi mapel Bahasa Inggris

- Total number of Support Vectors: 207
- Bias (offset): 6.849
- $w[x1] = 2.185$
- $w[x2] = 2.462$
- $w[x3] = 3.968$
- $w[x4] = 1.470$
- $w[x5] = 3.644$

Artinya : $[x1]$ = bobot nilai pengaruh dari $w[x1]$ terhadap Y sebesar 2.185

4.3. Prediksi Nilai UN Mapel Matematika

Tabel hasil 3. Hasil Percobaan SVM Nilai UN Mapel Matematika

<i>C</i> (Cost)	<i>k-fold</i>								
	3			5			10		
<i>KERNEL</i>	0,1	0,2	0,3	0,1	0,2	0,3	0,1	0,2	0,3
<i>Dot</i>	1.046	1.035	1.021	1.109	1.115	1.113	0.993	0.991	0.990
	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
<i>Radial</i>	0.209	0.207	0.209	0.348	0.364	0.365	0.322	0.318	0.317
	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
<i>Polynomial</i>	0.945	0.938	0.934	0.940	0.938	0.932	0.904	0.890	0.882
	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
	0.065	0.087	0.097	0.144	0.144	0.143	0.190	0.191	0.188
	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
	1.371	1.317	1.397	1.347	1.427	1.398	1.172	1.109	1.155
	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
	0.372	0.453	0.557	0.810	0.907	0.816	0.493	0.506	0.483
	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-

Dari eksperimen didapat RMSE paling kecil untuk mata pelajaran Matematika didapat dari *kernel type radial* dan nilai C(cost) 0,3 yaitu 0.882 +/- 0.188

Desain Prediksi mapel Matematika

Total number of Support Vectors: 207

Bias (offset): 6.281

- $w[X1] = 1.922$
- $w[X2] = 3.948$

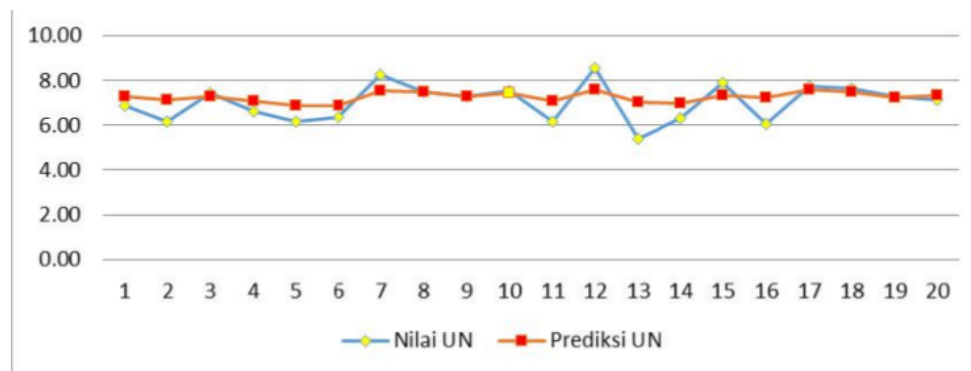
- $w[X3] = 6.701$
- $w[X4] = 8.783$
- $w[X5] = 7.557$

Artinya : $w[x1]$ = bobot nilai pengaruh dari $w[x1]$ terhadap Y sebesar 1.922

Tabel 4. RMSE terkecil dari hasil eksperimen untuk ke tiga Mata Pelajaran

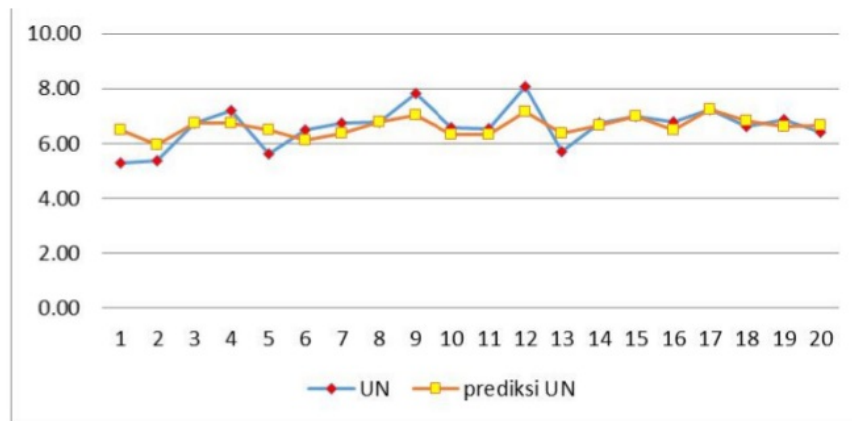
Mata Pelajaran	<i>k-fold</i>		<i>Kernel</i>	<i>RMSE</i>	W[X1]	W[X2]	W[X3]	W[X4]	W[X5]
Bahasa Indonesia	10	0,2	<i>Radial</i>	0.713 +/- 0.173	2.656	1.005	2.862	0.089	-0.623
Bahasa Inggris	10	0,3	<i>Radial</i>	0.586 +/- 0.066	2.185	2.462	3.968	1.470	3.644
Matematika	10	0,3	<i>Radial</i>	0.882 +/- 0.188	1.922	3.948	6.701	8.783	7.557

Dari hasil performa di atas dapat diketahui bahwa prediksi nilai Ujian Nasional untuk mata pelajaran Bahasa Indonesia dapat dilihat pada grafik berikut :



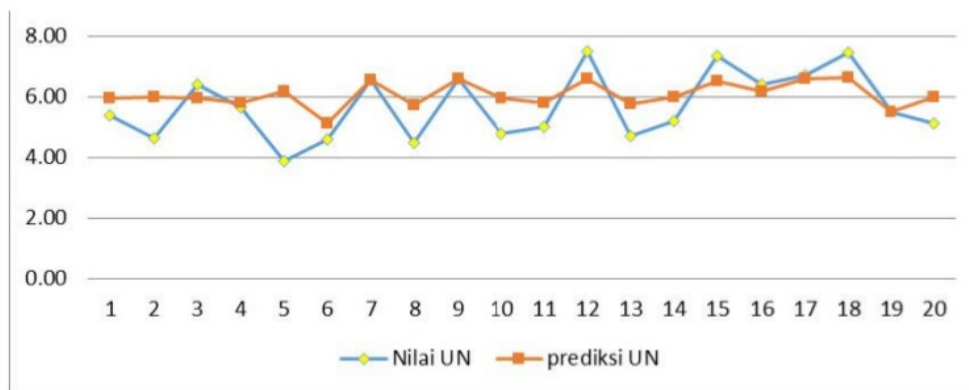
Gambar 3. Grafik Nilai Prediksi dan Nilai Ujian Nasional Mapel Bahasa Indonesia

Sedangkan prediksi nilai Ujian Nasional untuk mata pelajaran Bahasa Inggris dapat dilihat pada grafik berikut.



Gambar 4. Grafik Nilai Prediksi dan Nilai Ujian Nasional *Mapel Bahasa Inggris*

Sedangkan prediksi nilai Ujian Nasional untuk mata pelajaran Matematika dapat dilihat pada grafik berikut.



Gambar 5. Grafik Nilai Prediksi dan Nilai Ujian Nasional *Mapel Matematika*

Melihat kedekatan hasil ketiga grafik nilai prediksi permapel antara hasil nilai prediksi UN dan nilai UN maka model/performa tersebut bisa digunakan guna memprediksi nilai UN tahun pelajaran 2013/2014 untuk ketiga mata pelajaran (Bahasa Indonesia, Bahasa Inggris dan Matematika) dengan hasil prediksi dalam lampiran 3).

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Setelah melakukan percobaan dengan *input* data per-mata pelajaran yang di UN-kan, dengan menggunakan algoritma SVM maka didapatkan konfigurasi model performa untuk prediksi nilai Ujian Nasional sebagai berikut : *Root mean squared Error* (RMSE) terbaik untuk Bahasa Indonesia adalah 0.713 +/- 0.173, Bahasa Inggris sebesar 0.586 +/- 0.066, dan Matematika sebesar 0.882 +/- 0.188.

dengan konfigurasi prediksi menggunakan barometer. k -fold 10, C (cost) sebesar 0,1 dengan *type kernel radial* untuk mata pelajaran Bahasa Indonesia, k -fold 10, C (cost) sebesar 0,3 dengan *type kernel radial* untuk mata pelajaran Bahasa Inggris dan Matematika.

5.2. Saran-saran

Dalam penelitian ini telah menghasilkan prediksi yang optimal dan akurat, namun demikian untuk penelitian selanjutnya agar mendapatkan hasil yang lebih baik, perlu memperhatikan hal-hal sebagai berikut : Penelitian prediksi nilai UN Sekolah Menengah Kejuruan ini masih perlu dikembangkan, disamping menggunakan model performa konfigurasi *Support vektor machine*, dapat menambahkan atribut-atribut yang dapat mempengaruhi hasil nilai UN, misalnya : jenis kelamin, kondisi ekonomi siswa, dan pekerjaan orangtua. Untuk mendapatkan hasil prediksi yang lebih baik dapat dilakukan penelitian lebih lanjut dengan melakukan optimasi pada SVM dan dikomparasikan dengan algoritma yang lain.

PENUTUP

Syukur Alhamdulillah, berkat rahmat Allah SWT, penulis dapat menyelesaikan penyusunan tesis dengan judul : ALGORITMA SUPPORT VECTOR MACHINE UNTUK MEMREDIKSI NILAI UJIAN NASIONAL. Penyusunan tesis ini berhasil berkat adanya bantuan dan dukungan dari berbagai pihak. Penulis menyadari adanya keterbatasan penelitian ini, maka kritik, saran dan masukan yang membangun akan sangat membantu penulis dalam penelitian selanjutnya. Semoga tesis ini dapat bermanfaat untuk mengembangkan ilmu pengetahuan terutama dalam bidang sistem informasi.

PERNYATAAN ORIGINALITAS

"Saya menyatakan dan bertanggung jawab dengan sebenarnya bahwa Artikell ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya"
[Emi Rizky- P31.2011.01051]

DAFTAR PUSTAKA

- [1] Chun-Teck Lye, Lik-Neo Ng, Mohd Daud Hasan, Wei-wei Goh, Check-Yee Law, Noradzilah Ismail, *Predicting Pre-university Students' Mathematics Achievement*. ScienceDirect, 2010
- [2] Nuraeni, *Jaringan Syaraf Tiruan Untuk Mengukur Tingkat Korelasi Antara NEM Dengan IPK Kelulusan Siswa*. Jurnal Telkomnika Vol. 7 No. 3. Yogyakarta: tidak diterbitkan, Desember 2009
- [3] A. Hilda, *Komparasi metode Klasifikasi Data Mining untuk Prediksi Ketepatan Kelulusan Mahasiswa*, Thesis pada Program Pasca Sarjana Udinus, Semarang : tidak diterbitkan, 2012
- [4] Triswanto, *Komparasi Pemodelan Data Menggunakan Support Vector Machine dan Neural Network Untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa*. Thesis pada Universitas Dian Nuswantoro (DINUS) : tidak diterbitkan, 2013
- [5] Redho Pati, *Komputerisasi Iridologi Untuk Mendeteksi Kondisi Ginjal Menggunakan SUPPORT VECTOR MACHINE (SVM)*, Tugas Akhir pada Institut Teknologi Telkom Bandung: tidak diterbitkan, 2009
- [6] Sufandi, Unggul Utan, *Pengembangan sistem prediksi kemajuan belajar mahasiswa berbasis jaringan saraf tiruan: kasus universitas terbuka*. Thesis pada Institut Pertanian Bogor (IPB), Bogor : tidak diterbitkan, 2007
- [7] Iin Ernawati, *Prediksi Status Keaktifan Studi Mahasiswa dengan Algoritma C5.0 dan k-Nearest Neighbor*. Thesis pada Institut Pertanian Bogor (IPB), Bogor : tidak diterbitkan, 2008
- [8] Baha Sen & Emine Ucar. *Predicting and analyzing secondary education placement-test scores: A data mining approach*. Journal Expert Systems with Applications: An International Journal

archive – Volume 39 Issue 10, Page 9468-9478, Pergamon Press, Inc. Tarrytown, NY, USA, August 2012

[9] Peraturan Pemerintah Nomor 19 tentang Standar Nasional Pendidikan, Tahun 2005

[10] Nugroho, *SVM: Paradigma baru dalam Softcomputing dan aplikasinya*, Konferensi Nasional Sistem & Informatika, Bali, 15 November 2008

ALGORITMA SUPPORT VECTOR MACHINE

ORIGINALITY REPORT

21 %

SIMILARITY INDEX

17 %

INTERNET SOURCES

7 %

PUBLICATIONS

14 %

STUDENT PAPERS

PRIMARY SOURCES

1

www.scribd.com

Internet Source

3 %

2

pt.scribd.com

Internet Source

2 %

3

Submitted to Universitas Brawijaya

Student Paper

2 %

4

digilib.its.ac.id

Internet Source

1 %

5

es.scribd.com

Internet Source

1 %

6

Ernesto Pathros Ibarra Garcia, Pablo Medina Mora. "Model Prediction of Academic Performance for First Year Students", 2011 10th Mexican International Conference on Artificial Intelligence, 2011

Publication

1 %

7

kristian-nuki15.blogspot.com

Internet Source

1 %

Submitted to Universitas Negeri Jakarta

8

Student Paper

1%

9

Baha Şen, Emine Uçar, Dursun Delen.
"Predicting and analyzing secondary education
placement-test scores: A data mining
approach", Expert Systems with Applications,
2012

Publication

1%

10

media.neliti.com

Internet Source

1%

11

eprints.uny.ac.id

Internet Source

1%

12

text-id.123dok.com

Internet Source

1%

13

Submitted to Universitas Muria Kudus

Student Paper

1%

14

untirta.ac.id

Internet Source

1%

15

openlibrary.telkomuniversity.ac.id

Internet Source

1%

16

dwiaryaniwuryanto.wordpress.com

Internet Source

1%

17

Henny Dwi Bhakti. "Aplikasi Artificial Neural
Network (ANN) untuk Memprediksi Masa Studi

1%

Mahasiswa Program Studi Teknik Informatika
Universitas Muhammadiyah Gresik", Eksplora
Informatika, 2019

Publication

18	belajarsendri.blogspot.com Internet Source	1%
19	www.essays.se Internet Source	1%
20	pujiadilpmpjateng.wordpress.com Internet Source	<1%
21	smp3saketi.blogspot.com Internet Source	<1%
22	mafiadoc.com Internet Source	<1%
23	repository.usd.ac.id Internet Source	<1%
24	Submitted to Udayana University Student Paper	<1%
25	sawali.info Internet Source	<1%
26	Submitted to President University Student Paper	<1%
27	Muhamad Subhan Efendi, Helmie Arif Wibawa. "Prediksi Penyakit Diabetes Menggunakan	<1%

Algoritma ID3 dengan Pemilihan Atribut Terbaik", JUITA : Jurnal Informatika, 2018

Publication

Exclude quotes Off

Exclude matches < 5 words

Exclude bibliography On