

Perbandingan *Simple Logistic Classifier* dengan *Support Vector Machine* dalam Memprediksi Kemenangan Atlet

Ednawati Rainarli¹⁾, Arif Romadhan²⁾

¹⁾²⁾*Teknik Informatika, Teknik dan Ilmu Komputer, Universitas Komputer Indonesia
Jl. Dipati Ukur 112-114, Bandung*

¹⁾ednawati.rainarli@email.unikom.ac.id

²⁾arifromadhan19@gmail.com

Abstrak— Prediksi kemenangan atlet adalah hal yang harus dilakukan oleh pelatih ketika memutuskan pemain yang akan diturunkan dalam suatu pertandingan. Banyaknya faktor-faktor yang mempengaruhi kemenangan atlet membuat keputusan tersebut tidak mudah untuk ditentukan. Dalam penelitian ini akan dilakukan perbandingan dari penggunaan metode *Simple Logistic Classifier* (SLC) dengan *Support Vector Machine* (SVM) dalam memprediksi kemenangan atlet berdasarkan data kesehatan dan data latihan fisik. Data yang digunakan diambil dari 28 cabang olahraga perorangan. Rata-rata akurasi SLC dan SVM masing-masing diperoleh sebesar 80% dan 88%, sedangkan rata-rata kecepatan pemrosesan metode SLC dan SVM adalah 1,6 detik dan 0,2 detik. Hal ini menunjukkan bahwa penggunaan metode SVM lebih unggul daripada SLC, baik dari segi kecepatan maupun dari nilai akurasi yang dihasilkan. Selain pengujian akurasi, dilakukan pula pengujian terhadap 24 fitur yang digunakan dalam proses klasifikasi. Hasilnya diketahui bahwa pengurangan fitur melalui tahap seleksi mengakibatkan penurunan nilai akurasi. Berdasarkan hal tersebut disimpulkan bahwa semua fitur yang digunakan dalam penelitian ini adalah fitur yang berpengaruh dalam penentuan prediksi kemenangan atlet.

Kata Kunci— *Prediksi, Simple Logistic Classifier, Sports Data Mining, Support Vector Machine*

Abstract— A coach must be able to select which athlete has a good prospect of winning a game. There are a lot of aspects which influence the athlete in winning a game, so it's not easy by coach to decide it. This research would compare Simple Logistic Classifier (SLC) and Support Vector Machine (SVM) usage applied to predict winning game of athlete based on health and physical condition record. The data get from 28 sports. The accuracy of SLC and SVM are 80% and 88% meanwhile processing times of SLC and SVM method are 1.6 seconds dan 0.2 seconds. The result shows the SVM usage superior to the SLC both of speed process and the value of accuracy. There were also testing of 24 features used in the classifications process. Based on the test, features selection process can cause decreasing the accuracy value. This result concludes that all features used in this research influence the determination of a victory athletes prediction.

Keywords— *Prediction, Simple Logistic Classifier, Sports Data Mining, Support Vector Machine*

Article history:

Received 5 July 2017; Received in revised form 13 August 2017; Accepted 16 August 2017; Available online 28 October 2017

I. PENDAHULUAN

Secara sederhana prediksi kemenangan seorang atlet dapat ditentukan berdasarkan jumlah kemenangan yang tinggi atau berdasarkan pada prestasi yang telah diraihnya. Akan tetapi, tidak dapat dipungkiri bahwa ada banyak faktor lain yang dapat mempengaruhi kemenangan seorang atlet. Faktor kondisi kesehatan dan faktor kondisi fisik adalah faktor-faktor yang sering menjadi perhatian pelatih ketika akan menurunkan pemain dalam suatu pertandingan. Berkembangnya pemanfaatan *data mining* dalam bidang olah raga membuat prediksi kemenangan dapat dilakukan menggunakan teknik-teknik *data mining*. Selain untuk memprediksi kemenangan, *data mining* pun dapat dimanfaatkan untuk melihat performansi atlet, menentukan strategi dalam bertanding, ataupun mengidentifikasi bakat yang dimiliki oleh seorang atlet. Penelitian (Eggels, 2016) memanfaatkan teknik-teknik *data mining* untuk

menghasilkan prediksi gol yang mungkin terjadi dari sebuah susunan pemain yang diturunkan dalam pertandingan sepak bola. Kualitas dari gol yang dihasilkan diukur dengan menggunakan nilai probabilitas terjadinya gol. Penelitian ini berhasil menunjukkan bahwa dengan model yang dihasilkan mampu memprediksi lebih dari 50% kemenangan yang terjadi. *Data mining* juga dapat digunakan untuk menentukan prediksi kemenangan dari tim *football* (Leung & Joseph, 2014). Selain itu, (Solieman, 2006) juga memaparkan penerapan-penerapan *data mining* yang telah dilakukan dalam bidang olah raga. Dari penelitian tersebut disampaikan bahwa perlu dibangun sebuah repositori dari data-data histori pertandingan, performansi pemain baik berupa video, audio, maupun file data untuk digunakan dalam penelitian *data mining*.

(Haghighat, Rastegari, & Nourafza, 2013) membuat *literature review* yang membahas tentang penggunaan teknik *data mining* pada beberapa

cabang olah raga seperti basket, *rugby*, *soccer*, dan sepak bola. Dalam *literature review* ini dijelaskan beberapa metode *machine learning* yang digunakan, seperti metode Jaringan Syaraf Tiruan, Pohon Keputusan, metode Bayes, Regresi Logistik, metode Fuzzy mau pun dengan metode *Support Vector Machine* (SVM). Selain memaparkan metode - metode yang telah digunakan untuk memprediksi hasil pertandingan, disampaikan pula cara-cara pengumpulan data, seleksi fitur yang dilakukan oleh penelitian sebelumnya. Jumlah data set yang banyak serta pemilihan fitur yang relevan menjadi hal yang perlu diperhatikan untuk mendapatkan akurasi dari prediksi kemenangan pertandingan. Penelitian tersebut menyatakan bahwa nilai akurasi dari hasil - hasil penelitian tersebut masih rendah (Haghighat, Rastegari, & Nourafza, 2013). Hal ini menunjukkan bahwa penelitian - penelitian yang berkaitan dengan prediksi kemenangan atlet masih dapat terus dikembangkan dan dilakukan.

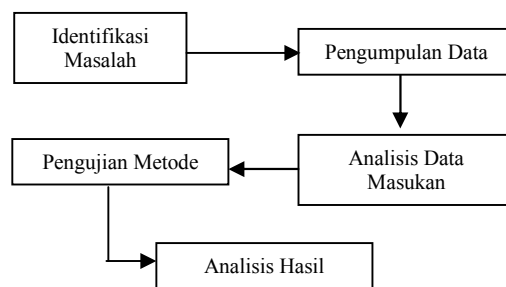
Penelitian oleh (Soto Valero, 2016) memanfaatkan *datamining* untuk memprediksi kemenangan tim dari permainan *baseball*. Penelitian ini menyimpulkan bahwa penggunaan metode klasifikasi khususnya *Support Vector Machine* (SVM) untuk memprediksi kemenangan tim menghasilkan nilai akurasi yang lebih baik dibandingkan menggunakan metode regresi untuk peramalan yaitu sebesar 60 %. Metode SVM sendiri adalah metode pembelajaran mesin yang cukup populer dan banyak dimanfaatkan dalam berbagai bidang ilmu, diantaranya adalah pada bidang bioinformatika, teks mining, pengenalan wajah, gambar dan pengenalan suara. (Wang, 2005) mendokumentasikan beberapa penelitian yang berkaitan dengan penggunaan dan pengembangan dari metode SVM. (Cao, 2012) menggunakan metode *Simple Logistics Classifier* (SLC), ANN, SVM dan Naïve Bayes untuk memprediksi kemenangan dalam pertandingan basket *National Basketball Association* (NBA). Sebagai data masukan digunakan statistik pertandingan NBA dari musim sebelumnya untuk memprediksi kemenangan klub basket. Dari penelitian ini didapatkan bahwa metode SLC lebih baik dari pada SVM, dengan besar akurasi 67.62%.

Berdasarkan penjelasan diatas diketahui bahwa SVM adalah metode yang paling baik digunakan untuk prediksi kemenangan (Soto Valero, 2016), sedangkan SLC adalah metode yang lebih baik dibandingkan dengan SVM untuk melakukan prediksi (Cao, 2012). Oleh karena itu, dalam penelitian ini akan dibandingkan penggunaan metode klasifikasi *Support Vector Machine* (SVM) dengan metode *Simple Logistics Classifier* (SLC) dalam memprediksi kemenangan atlet. Tujuan dari penelitian ini adalah untuk menguji metode yang lebih cocok digunakan untuk prediksi kemenangan serta mengevaluasi penggunaan data kesehatan dan

data kondisi fisik atlet sebagai fitur dalam penentuan kemenangan atlet.

II. METODE PENELITIAN

Untuk menguji penggunaan metode SLC dan SVM, maka dalam penelitian ini akan dilakukan beberapa tahapan. Tahapan dari penelitian ini ditunjukkan pada Gambar 1. Adapun, langkah-langkah dalam penelitian ini diawali dengan identifikasi masalah dilanjutkan dengan pengumpulan data yang akan digunakan, analisis data masukan serta pengujian penggunaan metode dan diakhiri dengan analisis hasil.



Gambar 1. Alur Metode Penelitian

A. Identifikasi Masalah

Untuk mengidentifikasi masalah yang akan diselesaikan maka dilakukan wawancara dengan pelatih fisik atlet dan pelatih yang bertugas menganalisis kondisi atlet serta pihak Pusat Pengolahan Data (Pulahta) Komite Olah Raga Nasional (KONI) Jawa Barat. Adapun, tujuannya adalah untuk mengetahui kebutuhan melakukan prediksi kemenangan atlet dan menentukan kandidat fitur yang akan digunakan dalam proses klasifikasi.

Selanjutnya, dilakukan studi literatur untuk mencari penelitian-penelitian yang berkaitan dengan prediksi kemenangan atlet khususnya yang memanfaatkan *soft computing*. Berkaitan dengan literatur yang ditemukan (Cao, 2012), diketahui bahwa penggunaan SVM dan SLC tidak selalu menunjukkan bahwa SVM lebih baik dari SLC begitu pula sebaliknya (Soto Valero, 2016). Berdasarkan hal ini maka dirumuskan masalah yang akan diselesaikan.

B. Pengumpulan Data

Tahapan selanjutnya yang dilakukan adalah pengumpulan data atlet dari 29 cabang olah raga untuk kejuaraan perorangan. Data yang diambil merupakan data kondisi kesehatan dan kondisi fisik dari setiap atlet beserta keberhasilan atlet dalam menyumbangkan medali baik emas, perak ataupun perunggu dari kejuaraan terakhir yang diikuti oleh atlet tersebut. Dari pengumpulan data didapatkan 310 data atlet. Detail dari jumlah atlet dari tiap cabang olah raga ditunjukkan pada Tabel 1. Walaupun pada Tabel 1. Sebaran data yang digunakan untuk tiap cabang olah raga tidak

merata akan tetapi, data atlet yang mendapatkan medali dengan yang tidak mendapatkan medali jumlahnya sama banyak.

TABEL 1. SEBARAN DATA TIAP CABANG OLAH RAGA

Cabor	Jumlah	Cabor	Jumlah
Anggar	17	PencakSilat	21
Angkatberat	2	Renang	3
Angkatbesi	2	Selam	9
Atletik	21	Senam	1
Balap sepeda	10	Sepatu Roda	6
Catur	1	Ski Air	8
Dayung	16	Squash	8
Gulat	26	Taekwondo	21
Judo	18	TarungDerajat	23
Karate	21	TenisLapangan	3
Kempo	18	TenisMeja	1
Layar	13	TerbangLayang	5
Loncat Indah	3	Tinju	13
PanjatTebing	13	Wushu	17

C. Analisis Data Masukan

Tahapan berikutnya adalah menyusun dan menyeleksi data set. Setiap data memiliki 24 fitur yang digunakan dan merupakan gabungan antara faktor-faktor kondisi kesehatan atlet serta kondisi fisik atlet. Ada 17 fitur yang menggambarkan kondisi kesehatan atlet dan ada 7 fitur yang mewakili kondisi fisik atlet. Setiap data yang telah terkumpul diberi label kelas. Label positif +1 diberikan untuk atlet yang mendapatkan medali pada kejuaraan terakhir yang diikuti dan label -1 untuk atlet yang tidak meraih medali. Data disimpan dalam format CSV. Dengan menggunakan semua data, dilakukan seleksi fitur. Dalam penelitian ini dibandingkan tiga metode seleksi fitur yaitu metode korelasi, *gain ratio* dan *chisquare test* untuk mendapatkan perankingan dari fitur-fitur yang digunakan untuk kemudian dipilih fitur yang akan dipakai dalam proses pengujian. Menurut (Villacampa, 2015) proses seleksi fitur perlu digunakan dengan tujuan untuk mengurangi jumlah dimensi dari fitur yang digunakan. Penelitian tersebut juga menyebutkan bahwa fitur yang telah diseleksi terbukti dapat meningkatkan efektifitas dan efisiensi dalam penyelesaian masalah data mining. Adapun, ketiga metode seleksi fitur ini digunakan karena metode ini adalah metode yang sering digunakan dalam penelitian.

D. Pengujian Metode

Untuk kebutuhan pengujian digunakan aplikasi WEKA. Langkah-langkah penggunaan aplikasi WEKA dapat dilihat pada (Witten, Frank, & Hall, 2011). Ada dua skenario pengujian yang digunakan yaitu: tanpa menggunakan seleksi fitur dan dengan melakukan seleksi fitur terlebih dahulu. Metode *K-fold Cross Validation* digunakan untuk menguji kedua metode *Simple Logistic Classifier* dan SVM dengan nilai $K=10$ (Witten, Frank, & Hall, 2011)

E. Analisis Hasil

Berdasarkan hasil pengujian yang dilakukan maka akan ditarik kesimpulan metode mana yang lebih baik untuk diterapkan pada kasus ini. Selain itu, melalui seleksi fitur akan diketahui pula faktor-faktor fisik dan kesehatan apa saja yang paling cocok digunakan dalam penentuan prediksi kemenangan atlet.

III. HASIL DAN PEMBAHASAN

Berdasarkan langkah kerja yang disampaikan pada metodologi penelitian maka berikut ini adalah hasil dari perankingan fitur-fitur yang digunakan sebagai dasar pemilihan fitur (*feature selection*) dan hasil pengujian akurasi dan waktu dari kedua metode.

A. Hasil Perankingan Fitur

Fitur yang digunakan ada 24 fitur. Fitur dari kondisi kesehatan atlet adalah nilai-nilai dari haemoglobin (Hb), leukosit (Leu), hematokrit (HT), trombosit (Tr), batang (Bt), segmen (Sg), limosit (Lm), monosit (Mn), LED, glukosa darah puasa (GDP), kolesterol total (Chol), asam urat (Au), trigliserida (Tg), ureum (Ur), kreatinin (Kr), SGOT, SGPT, sedangkan fitur dari kondisi fisik atlet adalah sit and reach (SR), trunk lift (TL), sit up (SU), push up (PU), back lift (BL), hardle jump (HJ), VO2 max (VO2).

Dengan menggunakan metode korelasi, *gain ratio* dan *chisquare* diperoleh nilai korelasi antara fitur dengan kemenangan atlet. Dari nilai korelasi tersebut maka disusunlah ranking dari setiap fitur. Urutan ranking dari masing-masing metode seleksi fitur ditunjukkan pada Tabel 2. Berdasarkan Tabel 2. maka fitur yang diseleksi adalah SU, Mn, Bt, Lm, Sg, LED, dan Hb. Pengurangan fitur ini dilakukan secara bertahap, dimulai dengan pengurangan tiga, empat dan tujuh fitur. Pengurangan fitur dihentikan ketika nilai akurasi dari pengelompokan kurang dari 80%. Pemilihan jumlah fitur yang akan dieliminasi didasarkan pada fitur yang berada pada ranking paling rendah. Pada Tabel 2. walaupun fitur HJ berada pada peringkat 22 pada metode korelasi akan tetapi pada *gain ratio* dan *chisquare test* fitur HJ berada pada ranking 15 sehingga fitur ini tidak termasuk yang dieliminasi.

Seleksi dari metode *gain ratio* dan *chisquare* juga menunjukkan hasil yang mirip dalam perankingan fitur. Akan tetapi, pada metode korelasi ada beberapa fitur seperti Bt, LED, Chol, dan SR yang hasil rankingnya bertolak belakang dengan *gain ratio* dan *chisquare*. Hal ini mungkin terjadi dikarenakan hubungan antara nilai fitur dengan kelas bukan merupakan relasi linear. Dugaan ini diperkuat oleh pernyataan (Chandrashekar & Sahin, 2014) yang menyatakan bahwa metode korelasi efektif untuk data-data yang memiliki hubungan yang linear .

TABEL 2. RANGKING DARI FITUR

Ranking	Korelasi	Gain Ratio	Chisquare
1	Bt	GDP	VO2
2	LED	Chol	Chol
3	GDP	VO2	GDP
4	Au	SGPT	SGPT
5	Tg	Au	Tg
6	Ur	Ur	Au
7	SGPT	Tg	Ur
8	Tr	SGOT	SGOT
9	Kr	TL	TL
10	Leu	Tr	Tr
11	Chol	SR	SR
12	HT	HT	HT
13	SGOT	Kr	BL
14	Hb	BL	Kr
15	Sg	HJ	HJ
16	TL	Leu	PU
17	BL	PU	Leu
18	VO2	Hb	Hb
19	PU	LED	Sg
20	SR	Sg	LED
21	Lm	Lm	Lm
22	HJ	Bt	SU
23	Mn	Mn	Mn
24	SU	SU	Bt

B. Hasil Perbandingan SLC dengan SVM

Ada dua skenario yang digunakan untuk menguji kedua metode ini. Pertama adalah pengujian tanpa penggunaan seleksi fitur dimana 24 fitur yang ada digunakan untuk menentukan prediksi kemenangan atlet. Pada skenario kedua dilakukan seleksi fitur yang terdiri dari penggunaan 21 fitur, 20 fitur dan 17 fitur. Urutan fitur diseleksi didasarkan pada hasil perankingan fitur. Hasil pengujian dari dua skenario untuk kedua metode tersebut ditunjukkan pada Tabel 3. Pengujian ini dilakukan dengan menggunakan metode *10-fold Cross Validation*. Adapun, fungsi kernel yang dipilih adalah fungsi *Gaussian Radial Basis Function*. Hasil dari empat kali percobaan didapatkan rata-rata nilai akurasi untuk SLC adalah sebesar 80% dan untuk SVM didapatkan rata-rata akurasi sebesar 88%. Dari segi waktu pemrosesan data dalam membentuk model, diperoleh SVM memiliki kecepatan rata-rata 0.21 detik yang lebih baik dari pada SLC dengan rata-rata 1.58 detik .

Penggunaan metode SVM menunjukkan hasil yang lebih baik dari segi ketepatan maupun kecepatan pembuatan model. Hal ini sejalan dengan penelitian yang telah dilakukan oleh (Soto Valero, 2016) yang mendapatkan hasil penggunaan metode SVM lebih baik dari pada SLC . Begitu pula dengan penelitian (Öğüt, Aktaş, Alp, & Doğanay, 2009) yang menunjukkan bahwa SVM

juga lebih baik dalam mengenali manipulasi keuangan dibandingkan dengan *Neural Network* (Öğüt, Aktaş, Alp, & Doğanay, 2009).

TABEL 3. PERBANDINGAN LOGISTIC CLASSIFIER DENGAN SVM

Fitur	SLC		SVM	
	Akurasi (%)	Waktu (detik)	Akurasi (%)	Waktu (detik)
24	81,6129	1,76	88,7097	0,21
21	79,0323	1,54	88,0645	0,21
20	80,6452	1,46	88,0645	0,21
17	80,0000	1,55	87,4194	0,2

IV. KESIMPULAN

Penelitian ini akan menguji kemampuan dari metode SLC dan SVM dalam melakukan prediksi kemenangan atlet berdasarkan data kesehatan dan data fisik atlet. Hasil dari pengujian telah ditunjukkan bahwa penggunaan histori data kondisi fisik dan kondisi kesehatan atlet dapat digunakan untuk memprediksi kemenangan atlet. Prediksi yang paling baik dilakukan dengan menggunakan semua fitur kesehatan dan fisik atlet. Penggunaan SVM dapat menjadi pilihan untuk kasus ini, akan tetapi nilai akurasi dari prediksi kemenangan atlet juga perlu mempertimbangkan histori data masukan atlet. Artinya, data latih yang dimasukkan adalah data terbaru dari kondisi atlet. Untuk kedepannya penggunaan data atlet bisa lebih spesifik untuk cabang olahraga tertentu sehingga memungkinkan untuk menambah ataupun mengurangi fitur yang lebih spesifik untuk digunakan. Pemilihan metode seleksi fitur tidak hanya menggunakan konsep filtering tapi dapat pula digunakan konsep hibrid.

UCAPAN TERIMA KASIH

Terima kasih kepada Pusat Pengolahan Data Komite Olahraga Nasional Indonesia Provinsi Jawa Barat yang telah mengijinkan penggunaan data atlet untuk digunakan dalam penelitian ini.

DAFTAR PUSTAKA

Cao, C. (2012). *Sport Data Mining Technology Used in Basketball Outcome Prediction*. Dublin Institute of Technology. Dubin: Dublin Institute of Technology.

Chandrashekar, G., & Sahin, F. (2014). A Survey on Feature Selection Methods. *Computer and Electrical Engineering*, 40 (1), 16-28.

Eggels, H. P. (2016). *Expected Goals in Soccer : Explaining Match Results using Predictive Analytics*. Master Thesis, Eindhoven University of Technology, Eindhoven.

Haghighat, M., Rastegari, H., & Nourafza, N. (2013). A Review of Data Mining Techniques for Result Prediction in Sports. *Advances in Computer Science : an International Journal*, 2 (5), 7-12.

- Leung, C. K., & Joseph, K. W. (2014). Sports data mining: predicting results for the college football games. *18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES 2014*, 35, pp. 710 – 719. Kanada: Elsevier B.V.
- Öğüt, H., Aktaş, R., Alp, A., & Doğanay, M. M. (2009). Prediction of Financial Information Manipulation by Using Support Vector Machine and Probabilistic Neural Network. *Expert Systems with Applications*, 36 (3), 5419-5423.
- Soliman, O. (2006). *Data Mining in Sport : A Research Overview*. Master Project, Penn State University, Management Information System, Pennsylvania.
- Soto Valero, C. (2016). Predicting Win-Loss outcome in MBL regular season games - A comparative study using data mining methods. *International Journal of Computer Science in Sport*, 15 (2), 91 - 109.
- Villacampa, O. (2015). *Feature Selection and Classification Methods for Decision Making : A comparative Analysis*. PhD Dissertation, Nova Southeastern University, Information System, Florida.
- Wang, L. (2005). *Support Vector Machine: Theory and Application*. Berlin: Springer Berlin Heidelberg.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques* (3 ed.). Massachusetts: Morgan Kaufmann Publisher.