

Aplikasi *Bagging* Regresi Logistik Dalam Menentukan Peluang Kemenangan Pemain Dalam Suatu Pertandingan

(Studi Kasus: Game *Age Of Empire 2*)

Zulhanif, Bertho Tantular, Gungum Darmawan, Neneng Sunengsih

Departemen Statistika FMIPA Universitas Padjadjaran

email dzulhanif@yahoo.com

Abstrak— Regresi logistik merupakan analisis regresi yang melibatkan variabel bebas (X) dan variabel tidak bebas (Y), dimana variabel tidak bebasnya mempunyai nilai integer (bilangan bulat). Variabel X sebagai prediktor bernilai numerik atau integer bisa satu atau lebih variabel bebas. Dalam penelitian ini akan dibahas suatu model prediksi yang menggabungkan model *Bagging* dengan model Regresi Logistik yang dikenal sebagai model *Bagging* Regresi Logistik (BRL). Model BRL akan digunakan untuk menentukan nilai peluang menang (kode =1) dan kalah (kode =0) dalam suatu permainan *age of empire 2*. *Age of Empire* merupakan suatu game yang digagas oleh *microsoft* lebih dari sepuluh tahun yang lalu, tapi mempunyai penggemar yang cukup banyak. Game ini merupakan game strategi, Setiap pertandingan terbagi menjadi dua team (team 1 dan team 2), setiap team bisa 2, 3 atau maksimal 4 player. Setiap pemain mendapatkan suku (*civilization*) secara random dimana terdapat 18 suku. Dengan menggunakan BRL, setiap pemain dapat di tentukan peluang menang atau kalah berdasarkan suku dan banyaknya team.

Kata kunci: *Age of Empire 2*, *Regresi Logistik*, *Bagging*, *Bagging Regresi Logistik*

I. PENDAHULUAN

Bagging adalah metode untuk menghasilkan sebuah nilai prediksi dari hasil agregasi beberapa prediktor. Agregasi prediktor ini merupakan rata-rata agregasi pada kasus non pengklasifikasian dan merupakan modus dari prediktor pada kasus pengklasifikasian. Proses agregasi didapat dari sekumpulan data set yang diambil secara acak melalui metode *bootstrap*, adapun model *bagging* ini dapat dikombinasikan dengan model lain seperti model Regresi Logistik [1]. Regresi logistik (kadang disebut model logistik atau model *logit*), sendiri dalam statistika digunakan untuk prediksi probabilitas kejadian suatu peristiwa dengan mencocokkan data pada fungsi logit kurva logistik. Metode ini merupakan model linier umum yang digunakan untuk regresi binomial. Seperti analisis regresi pada umumnya, metode ini menggunakan beberapa variabel prediktor, baik numerik maupun kategori. Misalnya, probabilitas bahwa orang yang menderita serangan jantung pada waktu tertentu dapat diprediksi dari informasi usia, jenis kelamin, dan indeks massa tubuh. Regresi logistik juga digunakan secara luas pada bidang kedokteran dan ilmu sosial, maupun pemasaran seperti prediksi kecenderungan pelanggan untuk membeli suatu produk atau berhenti berlangganan. Para peneliti telah menggunakan Analisis regresi Logistik untuk suatu pertandingan atau *game*. Dalam suatu pertandingan hasil (Y) dapat berupa dua kategori yaitu kalah dan menang, atau bisa juga tiga (3) kategori yaitu kalah, menang dan remis seperti dalam pertandingan catur. Dalam penelitian ini Analisis Regresi logistik di aplikasikan untuk memprediksi peluang menang dan kalah dalam suatu permainan *Age Of Empire 2* [9]. Dalam suatu pertandingan fenomena kalah dan menang dapat dibuat kode 0= kalah dan 1 =menang. Sehingga variabel respon dari pertandingan ini adalah biner (dua

kategori). Variabel -variabel yang memungkinkan dalam memprediksi peluang menang dan kalah adalah *Score Skill* (nilai kemahiran dari seorang pemain), *number of partner* (banyaknya rekan satu tim), dan *civilization* (karakter/peradaban pasukan yang dimainkan). Pada proses perhitungan pada model Bagging ini akan menggunakan software R sebagai alat bantu analisisnya. Pada makalah ini dilakukan pemodelan klasifikasi bagging pada regresi logistik pada kasus pengklasifikasian peluang kemenangan dari suatu permainan game age of empire 2, dimana metode ini diharapkan dapat memberikan hasil dengan tingkat akurasi yang cukup lebih baik

II. METODE PENELITIAN/ LANDASAN TEORI

Bagging [2], merupakan salah satu metode ensemble yang bertujuan untuk memperbaiki *stability* dan *accuracy* dari suatu model pembelajaran. Andaikan terdapat suatu set data $L \{(y_n, x_n), n=1, \dots, N\}$ yang mana y merupakan suatu label klas atau respon numerik pada kasus model logistik nilai respon y didefinisikan

. Asumsikan bahwa dari sekumpulan set data L dapat dibangun suatu prosedur untuk memprediksi nilai y berdasarkan input vektor \mathbf{x} untuk memprediksi y oleh $\varphi(\mathbf{x}, L)$. Jika suatu set data $\{L_k\}$ yang terdiri atas N *independent* observasi *underlying distribution* L .

Maka $\{L_k\}$ dapat dipergunakan sebagai prediktor yang lebih baik dalam memprediksi y dengan menggunakan $\varphi(\mathbf{x}, L_k)$. Pada kasus y yang numerik $\varphi(\mathbf{x}, L)$ dapat diestimasi dengan merata-ratakan nilai $\varphi(\mathbf{x}, L)$ atas k yang dinotasikan sebagai $\varphi_A(\mathbf{x}) = E_L \varphi(\mathbf{x}, L)$ yang mana simbol A berarti agregasi. Sedangkan pada kasus y merupakan suatu klas, metode untuk membuat agregasi dari $\varphi(\mathbf{x}, L_k)$ dengan menggunakan voting. Prinsi dasar dari metode bagging ini dengan mengambil secara berulang *sample bootsrap* [5]. $\{L^{(B)}\}$ dari L dalam bentuk $\varphi(\mathbf{x}, L^{(B)})$. Untuk y yang numerik maka

$$\varphi_B(\mathbf{x}) = av_B \varphi(\mathbf{x}, L^{(B)}) \quad (1)$$

Dan menggunakan metode voting jika y adalah kategori

$$\varphi_B(\mathbf{x}) = vot_B \varphi(\mathbf{x}, L^{(B)}) \quad (2)$$

perubahan kecil pada L dapat mengakibatkan perubahan besar pada φ . Hal ini telah dibahas lebih mendalam oleh Breiman [3][4].

Misalkan untuk setiap (y, \mathbf{x}) pada L dapat diambil secara independent dari distribusi peluang P . Andaikan y adalah nilai numerik dan $\varphi(\mathbf{x}, L)$ sebagai prediktor. Agregat prediktor pada L dapat dinyatakan sbg :

$$\phi_A(\mathbf{x}) = E_L \phi(\mathbf{x}, L) \quad (3)$$

Dengan mengambil \mathbf{x} sebagai nilai input dan y sebagai output maka

$$E_L (y - \phi(\mathbf{x}, L))^2 = y - 2y E_L \phi(\mathbf{x}, L) + E_L \phi^2(\mathbf{x}, L) \quad (4)$$

Dengan pemisalan $E_L \phi(\mathbf{x}, L) = \phi_A(\mathbf{x})$ serta menggunakan kesamaan $E Z^2 \geq (EZ)^2$ maka persamaan 4 dapat menjadi

$$E_L(y - \phi(\mathbf{x}, L))^2 \geq (y - \phi_A(\mathbf{x}))^2 \quad (5)$$

Pada persamaan 5 didapat bahwa MSE (*mean square error*) $\phi_A(\mathbf{x})$ akan lebih kecil jika dibandingkan MSE $\phi(\mathbf{x}, L)$ pada L , Seberapa besar perbedaan tersebut bergantung pada

$$|E_L \phi(\mathbf{x}, L)|^2 \leq E_L \phi^2(\mathbf{x}, L) \quad (6)$$

Pada persamaan 6 pengaruh ketidakstabilan terjadi jika $\phi(\mathbf{x}, L)$ tidak berubah secara signifikan yang menyebabkan agregasi yang dilakukan pada $\phi_A(\mathbf{x})$ tidak berpengaruh. Secara teori $\phi_A(\mathbf{x})$ selain bergantung pada nilai \mathbf{x} juga bergantung pada distribusi P pada L $\phi_A = \phi_A(\mathbf{x}, P)$. Sedangkan estimasi *bagging* dapat dinyatakan sbg:

$$\phi_B = \phi_B(\mathbf{x}, P_L) \quad (7)$$

yang mana P_L merupakan distribusi untuk setiap titik pengamatan $(y_n, \mathbf{x}_n) \in L$ (P_L disebut juga sebagai pendekatan bootstrap pada P).

Kualitas dari ϕ_B bergantung pada persamaan 7 jika tidak stabil maka estimator ϕ_B dapat memperbaiki tingkat akurasi dari nilai prediksi yang dilakukan, sebaliknya jika tidak stabil maka proses agregasi tidak terlalu berpengaruh dalam memperbaiki tingkat akurasi yang dibuat. Secara umum algoritma *bagging* [7][8][9] pada pengklasifikasian dapat ditulis sbb:

1. Lakukan secara berulang untuk $b = 1, 2, \dots, B$
 - a. Set replikasi bootstrap T_b dari data training T_n
 - b. Bangun fungsi klasifikasi $C_b(\mathbf{x}_i) = \{1, 2, \dots, k\}$ pada T_b
2. Kombinasikan fungsi $C_b(\mathbf{x}_i) = \{1, 2, \dots, k\}$ dengan menghitung voting mayoritas (modus),

$$C_f(\mathbf{x}_i) = \arg \max_j x_j \in y \sum_{b=1}^B I(C_b(\mathbf{x}_i) = j)$$

Faktor yang sangat penting dalam memperbaiki tingkat akurasi pada metode *bagging* ini adalah stabilitas pada waktu membangun ϕ . Unstabilitas dapat terjadi ketika

III. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini record hasil pertandingan *game Age of Empire 2*. Pertandingan dilakukan secara online melalui software Hamachi. Setiap pemain yang join ke dalam *game room* bersifat *independent* baik *civilization* maupun team di *setting* secara acak. Variabel yang dilibatkan dalam penelitian ini adalah $X_1 = \text{strategy}$ pemain, $X_2 = \text{Posision}$ (Posisi Pemain), $X_3 = \text{civilization}$ (Suku bangsa) serta variabel *dependent* nya adalah $Y = \text{Kalah-Menang}$.

Ukuran sampel sebanyak 412, yang terdiri atas permainan 4vs4, 3vs3 dan 2vs2. *Score* dari skill terbagi menjadi 3 yaitu cupu, menengah dan jenderal. Variabel X_2 terbagi menjadi 1 teman, 2 teman dan 3 teman. X_3 adalah suku bangsa (*civilization*) di *setting* secara acak untuk semua pemain. Setiap pemain mempunyai suku bangsa yang berbeda dalam satu game kode untuk suku bangsa di buat kode sebagai berikut; Azteks(1), Briton(2), Byzantyne(3), Celt(4), Chinesse(5), Frank(6), Goth(7), Japanese(8), Koreans(9), Huns(10), Mayans(11), Mongol(12), Persian(13), Saracens(14),

Spanish(15), *Teuton*(16), *Turky*(17), *Viking*(18). Selain itu player yang online untuk memainkan permainan bersifat saling independent.

Dalam proses pembentukan model klasifikasi dengan metode *Bagging* Regresi Logistik ini maka dilakukan serangkainya tahapan sbb:

1. Membagi data yang dipergunakan menjadi dua bagian yang terdiri atas data training dan data testing dengan perbandingan 75% dan 25%
2. Mengevaluasi besarnya kesalahan klasifikasi dari data training dan data testing
3. Membuat model prediksi

Dengan bantuan software didapat model klasifikasi dengan tingkat kepentingan variabel prediktor dan akurasi klasifikasi sbb:

A. Tingkat Kepentingan Variabel Prediktor

Tabel 1 Tingkat Kepentingan Variabel Prediktor

Variabel Prediktor	Nilai
civilizationaztecs	1.478103996
civilizationspain	1.305562769
civilizationkoreans	1.267481207
civilizationpersians	1.12895725
civilizationfranks	1.034412972
civilizationturks	1.014598495
civilizationsaracens	0.796341745
civilizationgoths	0.778968985
civilizationchina	0.771463888
civilizationjapanese	0.688964095
civilizationmongols	0.566278441
civilizationbritons	0.534462703
civilizationteutons	0.493210399
civilizationbyzantine	0.472417952
civilizationmayans	0.372220737
civilizationhuns	0.012216123
civilizationcelts	0.009613953
strategisemiboom	0.006454173
strategirush	0.006357929
strategiboom	0.006295706
pos	0.73463306

B. Tingkat Akurasi Klasifikasi

Tingkat akurasi klasifikasi ditinjau dari data training dan data testin sbb

B.1 Data Training

Tabel 2 Tabel Klasifikasi Data Training

	Menang	Kalah	Total
Menang	135	96	231
Kalah	73	108	181
Total	208	204	412

Dari data training pada tabel 2 dapat dilihat tingkat akurasi sebesar 58.98%

B.2 Data Testing

	Menang	Kalah	Total
Menang	60	25	85
Kalah	14	39	43
Total	74	64	138

Dari data training pada tabel dapat dilihat tingkat akurasi yang 71.73%. Sedangkan model klasifikasinya sendiri untuk membuat model prediksinya ada pada lampiran 1.

IV. SIMPULAN DAN SARAN

- Hasil analisis menunjukkan adanya kekurangan akuratan hasil klasifikasi pada data testing yang berpotensi menyebabkan *over fitting* dari model klasifikasi yang dibentuk.
- Pemodelan klasifikasi dengan metode ini perlu diuji lagi berkenaan dengan mempertimbangkan banyaknya *sample bootstrap*
- Pereduksian jumlah varabel prediktor menjadi hal yang dapat dipertimbangkan untuk mengurangi kesalahan dari model klasifikasi yang dibuat.

UCAPAN TERIMAKASIH

Terima kasih kami ucapkan kepada Ketua Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Padjadjaran dan Staf Fakultas matematika dan Ilmu Pengetahuan Alam Universitas Padjadjaran yang telah memberikan dana Hibah penelitian sehingga penelitian ini dapat berjalan dengan lancar.

DAFTAR PUSTAKA

- [1] Agresti, Alan. 2007. An Introduction to Categorical Data Analysis. New Jersey: John Wiley & Sons. Inc.
- [2] Bauer, E. and R. Kohavi. 1999. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Machine Learning 36: 105–139.
- [3] Breiman, L., J. Friedman, R. Olshen, and C. Stone. 1984. Classification and Regression Trees. Belmont, CA: Wadsworth.
- [4] Breiman, L. (1994). Heuristics of instability in model selection, Technical Report, Statistics Department, University of California at Berkeley (to appear, Annals of Statistics).
- [5] Diamond, Douglas W., 1984, Efron, B., & Tibshirani, R. 1993. "An Introduction to the Bootstrap". Chapman and Hall.
- [6] Freund, Y. and R. E. Schapire. 1997. A decision-theoretic generalization of online learning and an application to boosting. Journal of Computer and System Sciences 55(1): 119–139.
- [7] Friedman, J. 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics 29: 1189–1232.
- [8] Friedman, J., T. Hastie, and R. Tibshirani. 2000. Additive logistic regression: a statistical view of boosting. Annals of Statistics 28: 337–407.
- [9] Hastie, T., R. Tibshirani, and J. Friedman. 2001. The Elements of Statistical Learning New York: Springer.
- [10] Microsoft (1997). "Age of Empire 2 The Conqueror Expansion", Ensemble Studios