

Prediksi Kemenangan *eSport* DOTA 2 Berdasarkan Data Pertandingan

Eduardus Hardika Sandy Atmaja

Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Sanata Dharma
email: edo@usd.ac.id

Abstract

DOTA 2 is one of the eSports that are in great demand both by the general society and the game professional communities. They compete with each other to develop the best strategy to defeat all enemies they faced. In order to develop the best strategy, a good and accurate analysis system is needed. Data mining can be used to solve these problems by digging valuable information from dataset using certain method. Prediction method is one of the methods in data mining that is most appropriate for finding the winning predictions for the DOTA 2 game. One method that is quite simple and can be used is Naive Bayes. The results of this study indicate that Naive Bayes can make predictions well with an accuracy of 98,804 %. The data used in this research as much as 50000 that obtained from open data. It is expected that this research can assist players in providing information for developing game strategies.

Keywords — data mining, DOTA 2, eSports, prediction

1. Pendahuluan

eSports merupakan kepanjangan dari *electronic sport* atau olahraga elektronik yang menggunakan gim sebagai bidang yang dipertandingkan. Pertandingan dilakukan secara *online* dengan skema multipemain yang dapat diikuti oleh orang-orang dari belahan dunia manapun. Karena kemudahan aksesnya, *eSports* menjadi cabang olahraga baru yang banyak diminati oleh berbagai kalangan usia. Tingginya anemo masyarakat terhadap *eSport*, membuat berbagai perusahaan gim mengadakan kejuaraan baik dalam lingkup regional maupun internasional dengan hadiah yang fantastis. Berawal dari situlah, muncul tim-tim yang secara profesional menggeluti *eSport* dan menjadikannya pekerjaan dengan gaji tinggi.

DOTA 2 merupakan salah satu *eSports* yang sudah mendunia dan selalu mengadakan turnamen dengan hadiah yang sangat fantastis. Model dari permainan ini adalah pertandingan antara dua tim yang saling membunuh untuk mempertahankan area masing-masing. Di dalam permainan, dua tim yang saling beradu diberi nama tim *Radiant* dan tim *Dire*. Tim-tim yang mengikuti turnamen DOTA 2 juga terdiri dari orang-orang yang profesional. Selain mencari pemain yang handal, mereka juga melakukan analisis terhadap data-data gim yang diikuti guna mendukung penentuan strategi pertandingan. Proses analisis tersebut tentunya membutuhkan metode tertentu guna mendapatkan hasil yang akurat. Salah satu teknik yang dapat digunakan untuk menyelesaikan permasalahan analisis tersebut adalah *data mining*.

Data mining merupakan salah satu bidang kajian yang dapat menemukan informasi berharga dari sejumlah data yang besar [1]. Informasi yang dapat ditemukan dapat berupa pola atau keterkaitan data. Terdapat beberapa metode dalam *data mining* yang dapat digunakan untuk menggali informasi berharga, salah satunya adalah prediksi. Prediksi merupakan salah satu metode untuk memperkirakan sebuah nilai pada masa yang akan datang berdasarkan model klasifikasi yang diperoleh sebelumnya [2].

Metode prediksi dapat digunakan pada data pertandingan DOTA 2 untuk memprediksi kemenangan diakhir permainan nanti. Proses prediksi diawali dengan membuat sebuah model klasifikasi berdasarkan data pertandingan yang sudah ada untuk selanjutnya

digunakan oleh proses prediksi dalam menentukan hasil kemenangan. Dengan adanya prediksi kemenangan lebih awal maka tim dapat segera menyusun ulang strategi perlawanan ketika hasil prediksi menunjukkan kekalahan. Sehingga dapat meningkatkan kemungkinan kemenangan.

Metode prediksi dalam *data mining* sudah banyak diimplementasikan untuk menyelesaikan beberapa permasalahan, salah satunya adalah prediksi *rating* film dengan menggunakan metode *Naive Bayes* [3]. Tujuan dari penelitian tersebut adalah untuk membantu calon penonton mengetahui *rating* film yang belum mereka tonton berdasarkan beberapa informasi yang disediakan. Dengan mengetahui *rating* film maka penonton dapat memutuskan untuk menonton film tersebut atau tidak. Penelitian tersebut menghasilkan akurasi sebesar 55,80% dengan *precision* 32,41% dan *recall* 46,70%.

Penelitian lain juga telah berhasil mengimplementasikan *Naive Bayes* untuk melakukan prediksi harga emas [4]. Penelitian tersebut bertujuan untuk memprediksi harga emas berdasarkan informasi kondisi perekonomian, laju inflasi, penawaran dan permintaan serta masih banyak lagi. Dengan mengetahui harga emas lebih awal maka pelanggan dapat menentukan apakah akan membeli atau menjual emas tergantung pilihan mana yang lebih menguntungkan. Penelitian tersebut menghasilkan akurasi sebesar 75%.

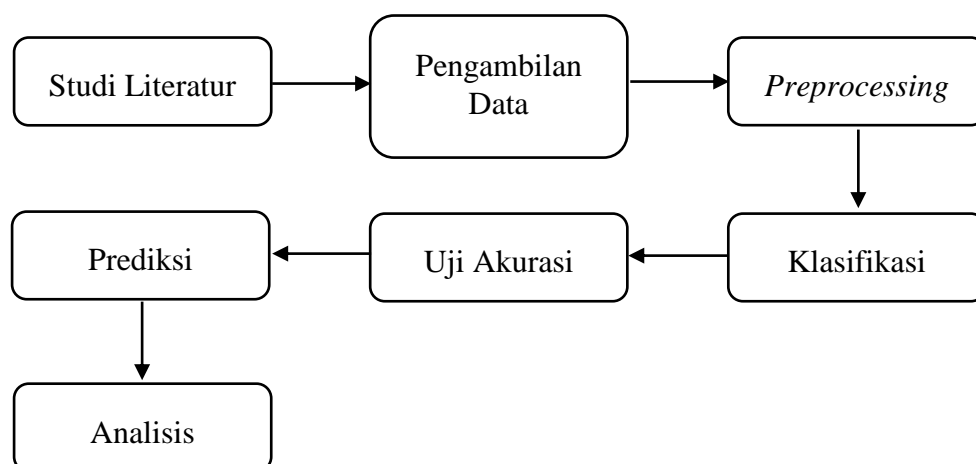
Naive Bayes juga telah diimplementasikan untuk melakukan prediksi kelulusan mahasiswa [5]. Penelitian tersebut bertujuan untuk memprediksi kelulusan mahasiswa untuk masuk ke perguruan tinggi berdasarkan informasi nilai bahasa indonesia, bahasa inggris, pengetahuan umum, matematika dasar dan wawancara. Hasil penelitian tersebut sangat baik dengan akurasi sebesar 94%.

Berdasarkan uraian di atas dan didukung oleh beberapa penelitian terdahulu maka penelitian ini mencoba untuk melakukan prediksi kemenangan *eSport* DOTA 2 berdasarkan data pertandingan dengan menggunakan *Naive Bayes*. Data pertandingan DOTA 2 didapatkan dari situs www.kaggle.com [6]. Dengan adanya penelitian ini, diharapkan dapat membantu pihak yang terkait untuk mengetahui prediksi kemenangan tim sejak dini sehingga dapat membantu tim dalam menyusun strategi pertandingan.

2. Metodologi

2.1 Tahap penelitian

Beberapa tahap penelitian telah dilalui untuk menyelesaikan penelitian ini yang digambarkan pada Gambar 1 di bawah ini.



Gambar 1. Tahap Penelitian

Gambar 1 merupakan rangkaian dari tahap penelitian yang dilalui pada penelitian ini.

Studi literatur merupakan tahap pertama yang dilakukan yaitu kegiatan untuk mempelajari teori yang berkaitan dengan penelitian ini termasuk di dalamnya adalah mempelajari penelitian terdahulu guna mendukung penelitian ini. Beberapa literatur yang dipergunakan antara lain *website*, prosiding, jurnal, buku dan publikasi ilmiah. Pengambilan data dilakukan setelah teori-teori selesai dipelajari. Data diambil dari situs *www.kaggle.com* [6] berupa data pertandingan DOTA 2 yang terdiri dari 8 atribut. Tahap ketiga adalah *preprocessing* yaitu melakukan seleksi data dan pembersihan data untuk menghilangkan data yang tidak relevan supaya siap untuk diolah. Tahap ke empat adalah klasifikasi yaitu melakukan proses perhitungan *Naive Bayes* untuk menemukan model klasifikasi. Setelah model didapatkan, selanjutnya menguji akurasi dari model yang didapatkan. Dari model yang telah didapatkan selanjutnya dapat dilakukan prediksi berdasarkan data tunggal yang dimasukkan. Langkah terakhir adalah melakukan analisa terhadap hasil yang telah diperoleh

2.2 Data

Data dalam penelitian ini terdiri dari 8 atribut yang dijelaskan sebagai berikut.

- match_id*, merupakan nomor urut pertandingan.
- duration*, merupakan durasi pertandingan.
- tower_status_radiant*, *hit point* tower milik tim *radiant*.
- tower_status_dire*, *hit point* tower milik tim *dire*.
- barracks_status_dire*, *hit point* barak milik tim *dire*.
- barracks_status_radiant*, *hit point* barak milik tim *radiant*.
- first_blood_time*, waktu terjadi kematian pertama.
- radiant_win*, status kemenangan *radiant*. Nilai *true* jika *radiant* menang dan nilai *false* jika *radiant* kalah.

Sampel data yang digunakan untuk percobaan sebanyak 15 data dengan pengujian 3-*fold cross validation*. Data 1 sampai 10 akan menjadi data latih sedangkan data 11 sampai 15 akan menjadi data uji. Label klasifikasi yang digunakan adalah atribut *radiant_win*. Sampel data diperlihatkan pada Tabel 1.

Tabel 1. Sampel data

<i>match_id</i>	<i>duration</i>	<i>tower_status_radiant</i>	<i>tower_status_dire</i>	<i>barracks_status_dire</i>	<i>barracks_status_radiant</i>	<i>first_blood_time</i>	<i>radiant_win</i>
1	2375	1982	4	3	63	1	TRUE
2	2716	256	1972	63	48	190	FALSE
3	3085	4	1924	51	3	40	FALSE
4	3830	548	384	48	15	176	TRUE
5	1562	1983	262	51	63	143	TRUE
6	3256	1792	260	51	52	73	TRUE
7	1728	2047	256	48	63	27	TRUE
8	2820	4	1572	63	3	168	FALSE
9	2754	1852	256	48	63	4	TRUE
10	3388	4	1796	63	51	80	FALSE
11	2798	1982	256	48	63	102	TRUE
12	3098	1796	260	51	63	228	TRUE
13	3355	6	1572	15	3	17	FALSE
14	2295	256	1588	63	48	55	FALSE

15	1331	1926	2047	63	63	123	FALSE
----	------	------	------	----	----	-----	-------

2.3 Naive Bayes

Metode *naive bayes* berfungsi untuk menemukan model klasifikasi dengan cara menghitung probabilitas keanggotaan dari suatu kelas. Secara umum rumus dari metode *Naive Bayes* diperlihatkan pada Persamaan 1 [2]. Keterangan rumus adalah sebagai berikut: x adalah data yang belum diketahui kelasnya, c adalah hipotesis x yang merupakan suatu kelas spesifik, $P(c|x)$ adalah probabilitas c berdasarkan x , $P(c)$ adalah probabilitas hipotesis x

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

Naive bayes biasa digunakan untuk data bertipe kategorikal namun dapat juga digunakan untuk data kontinu dengan perlakuan khusus yaitu dengan menghitung distribusi gaussian. Rumus distribusi gaussian diperlihatkan pada Persamaan 2 [7]. Keterangan rumus adalah sebagai berikut: μ adalah rata-rata data, σ adalah standar deviasi.

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi}(\sigma)} e^{-\frac{(x_i - \mu)^2}{2(\sigma)^2}} \quad (2)$$

Untuk mengetahui keberhasilan dari model yang didapatkan maka perlu dilakukan uji akurasi. Adapun rumus uji akurasi diperlihatkan pada Persamaan 3 [7].

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi benar}}{\text{Jumlah prediksi yang dilakukan}} \quad (3)$$

3. Hasil Penelitian dan Analisis

3.1 Perhitungan probabilitas kelas latih

Terdapat dua kelas hasil klasifikasi yaitu *TRUE* dan *FALSE* pada atribut *radiant_win* berdasarkan Tabel 1. Maka dicari probabilitas setiap kelas dengan membagi total kemunculan kelas c dengan total keseluruhan kelas.

Tabel 2. Probabilitas kelas

Jumlah Kelas		Probabilitas Kelas	
TRUE	FALSE	TRUE	FALSE
6	4	0,6	0,4

3.2 Perhitungan *mean* dan standar deviasi untuk membuat model klasifikasi

Data yang ada pada Tabel 1 adalah data kontinu maka perhitungan probabilitas dilakukan dengan menghitung distribusi gaussian. Namun sebelum itu perlu di hitung mean dan standar deviasi untuk setiap atribut berdasarkan kelas latih. Hasil dari perhitungan mean dan standar deviasi di perlihatkan pada Tabel 3, Tabel 4, Tabel 5, Tabel 6, Tabel 7, Tabel 8.

Tabel 3. Hasil *mean* dan standar deviasi atribut *duration*

<i>duration</i>	TRUE	FALSE
<i>Mean</i>	2584,2	3002,25
<i>STDEV</i>	801,45	260,1945186

Tabel 4. Hasil *mean* dan standar deviasi atribut *tower_status_radiant*

<i>tower_status_radiant</i>	TRUE	FALSE
<i>Mean</i>	1700,7	67
STDEV	522,6	109,1192009

Tabel 5. Hasil *mean* dan standar deviasi atribut *tower_status_dire*

<i>tower_status_dire</i>	TRUE	FALSE
<i>Mean</i>	237	1816
STDEV	113,85	154,8676855

Tabel 6. Hasil *mean* dan standar deviasi atribut *barracks_status_dire*

<i>barracks_status_dire</i>	TRUE	FALSE
<i>Mean</i>	41,5	60
STDEV	17,27	5,196152423

Tabel 7. Hasil *mean* dan standar deviasi atribut *barracks_status_radiant*

<i>barracks_status_radiant</i>	TRUE	FALSE
<i>Mean</i>	53,167	26,25
STDEV	17,535	23,27418097

Tabel 8. Hasil *mean* dan standar deviasi atribut *first_blood_time*

<i>first_blood_time</i>	TRUE	FALSE
<i>Mean</i>	70,667	119,5
STDEV	67,746	61,65022303

3.3 Perhitungan distribusi gaussian

Setelah *mean* dan standar deviasi (model klasifikasi) untuk setiap atribut didapatkan, selanjutnya dilakukan perhitungan distribusi gaussian berdasarkan Persamaan 2. Pada tahap ini data uji digunakan untuk mengetahui kesesuaian klasifikasi hasil perhitungan dengan klasifikasi pada data asli. Sebagai contoh diambil data dengan *match_id* = 14 maka didapatkan seperti yang diperlihatkan pada Tabel 9.

Tabel 9. Hasil distribusi gaussian untuk *match_id* = 14

Atribut	TRUE	FALSE
<i>duration</i>	0,013207312	0,000615182
<i>tower_status_radiant</i>	0,000382439	0,008523694
<i>tower_status_dire</i>	9,92585E-33	0,010848884
<i>barracks_status_dire</i>	0,044240919	0,148182468
<i>barracks_status_radiant</i>	0,091246495	0,053449736
<i>first_blood_time</i>	0,04720247	0,02940129

3.4 Probabilitas akhir setiap kelas

Pada tahap ini dilakukan perkalian untuk semua probabilitas pada Tabel 2 dan Tabel 9 sesuai dengan kelasnya masing-masing.

$$\begin{aligned} P(x|\text{TRUE}) &= 0,6 * 0,013207312 * 0,000382439 * 9,92585E-33 * 0,044240919 * \\ & 0,091246495 * 0,04720247 \\ &= 5,73192E-42 \end{aligned}$$

$$\begin{aligned} P(x|\text{FALSE}) &= 0,6 * 0,000615182 * 0,008523694 * 0,010848884 * 0,148182468 * \\ & 0,053449736 * 0,02940129 \\ &= 5,29889E-12 \end{aligned}$$

3.5 Normalisasi

Setelah ditemukan probabilitas akhir selanjutnya dilakukan normalisasi untuk mendapatkan nilai yang mendekati atau sama dengan satu yang menunjukkan hasil klasifikasi.

$$\begin{aligned} P(\text{TRUE}) &= 5,73192E-42 / (5,73192E-42 + 5,29889E-12) \\ &= 1,08172E-30 \end{aligned}$$

$$\begin{aligned} P(\text{FALSE}) &= 5,73192E-42 / (5,73192E-42 + 5,29889E-12) \\ &= 1 \end{aligned}$$

Nilai hasil normalisasi untuk data $match_id = 14$ yang mendekati atau sama dengan 1 adalah *FALSE* maka dapat disimpulkan $match_id = 14$ dikategorikan sebagai *FALSE* atau tim *radiant* kalah dari tim *dire*. Tabel 10 menunjukkan hasil klasifikasi untuk semua data uji.

Tabel 10. Hasil klasifikasi

<i>match_</i> <i>id</i>	<i>duration</i>	<i>tower_</i> <i>status_</i> <i>radiant</i>	<i>tower_</i> <i>status_</i> <i>dire</i>	<i>barracks_</i> <i>status_</i> <i>dire</i>	<i>barracks_</i> <i>status_</i> <i>radiant</i>	<i>first_</i> <i>blood_</i> <i>time</i>	<i>radiant_</i> <i>win</i> <i>(data)</i>	<i>radiant_</i> <i>win</i> <i>(uji)</i>
11	2798	1982	256	48	63	102	TRUE	TRUE
12	3098	1796	260	51	63	228	TRUE	TRUE
13	3355	6	1572	15	3	17	FALSE	FALSE
14	2295	256	1588	63	48	55	FALSE	FALSE
15	1331	1926	2047	63	63	123	FALSE	TRUE

3.6 Akurasi

Untuk menghitung akurasi digunakan *confusion matrix*, yaitu tabel yang menggambarkan data yang terprediksi benar dan data yang terprediksi salah. Berdasarkan Tabel 10 data $match_id$ 11, 12, 13 dan 14 diklasifikasikan dengan benar sedangkan data $match_id$ 15 tidak diklasifikasikan dengan benar. Maka tabel *confusion matrix* diperlihatkan pada Tabel 11.

Tabel 11. *Confusion Matrix* data sampel

		Hasil Prediksi	
		TRUE	FALSE
Data Asli	TRUE	2	0
	FALSE	1	2

Berdasarkan Tabel 11 dan mengacu pada Persamaan 3 maka akurasi dapat dihitung sebagai berikut.

$$\text{Akurasi} = (4 / 5) * 100 \% = 80 \%$$

Berdasarkan perhitungan di atas, hasil akurasi *naive bayes* untuk 15 data pertandingan game DOTA 2 cukup baik yaitu sebesar 80%.

3.7 Pengujian keseluruhan data

Selanjutnya dilakukan pengujian terhadap keseluruhan data dengan menggunakan *tools Weka* [8]. Total data yang digunakan adalah 5000 [6] dengan pengujian *3-fold cross validation*, menghasilkan *confusion matrix* diperlihatkan pada Tabel 12.

Tabel 12. *Confusion Matrix* keseluruhan data

		Hasil Prediksi	
		TRUE	FALSE
Data Asli	TRUE	25618	325
	FALSE	273	23784

Berdasarkan Tabel 12 dan mengacu pada Persamaan 3 maka akurasi dapat dihitung sebagai berikut.

$$\text{Akurasi} = (49402 / 50000) * 100 \% = 98,804 \%$$

Berdasarkan perhitungan di atas, hasil akurasi *naive bayes* untuk keseluruhan data pertandingan game DOTA 2 sangat baik yaitu sebesar 98,804 %.

4. Kesimpulan

Berdasarkan serangkaian kegiatan yang telah dilakukan maka dapat disimpulkan bahwa algoritma *Naive Bayes* dapat digunakan untuk melakukan prediksi dengan baik. Akurasi yang didapatkan sebesar 98,804 % dengan *3-fold cross validation* dengan jumlah data 50000. Data yang digunakan dalam penelitian ini adalah data pertandingan saja, penelitian selanjutnya dapat mencoba mengkombinasikan antara data pertandingan dengan data lain seperti *player rating*, *item purchased*, *ability used* dan lainnya sehingga dihasilkan informasi yang lebih detail.

Referensi

- [1] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* 3rd Edition. Morgan Kaufmann.
- [2] Kusriani dan Luthfi, E. T. (2009). *Algoritma Data Mining*. Andi.
- [3] Pratiwi, R. W., & Nugroho, Y. S. (2016). Prediksi Rating Film Menggunakan Metode Naïve Bayes. *Jurnal Teknik Elektro*, 8(2), 59-63.
- [4] Guntur, M., Santony, J., & Yuhandri. (2018). Prediksi Harga Emas dengan Menggunakan Metode Naïve Bayes dalam Investasi untuk Meminimalisasi Resiko. *Jurnal RESTI : Rekayasa Sistem dan Teknologi Informasi*, 2(1), 354-360.
- [5] Syarli & Muin, A. A., (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmiah Ilmu Komputer*, 2(1), 22-26.
- [6] Anzelmo, D. (Diperbaharui pada November 2019). Dota 2 Matches. <https://www.kaggle.com/devinanzelmo/dota-2-matches>. Diakses pada tanggal 3 Januari 2020.
- [7] Prasetyo, E. (2012). *Data Mining: Konsep dan Aplikasi menggunakan Matlab*. Andi.
- [8] Weka versi 3.8.4. Diunduh dari <https://www.cs.waikato.ac.nz/~ml/weka/>.