

# ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI PREDIKSI KEMENANGAN DALAM GIM DOTA 2 MENGGUNAKAN LOGISTIC REGRESSION

Muhamad Dio Damiyati 1) , Yuda Kurnia Nurul Fikri 2)  
Jurusan Teknik Informatika Fakultas Sains dan Teknologi  
Universitas Islam Negeri Syarif Hidayatullah Jakarta

<sup>1</sup>[muhamad.dio18@mhs.uinjkt.ac.id](mailto:muhamad.dio18@mhs.uinjkt.ac.id)

<sup>2</sup>[yudakurnia.nurulfikri18@mhs.uinjkt.ac.id](mailto:yudakurnia.nurulfikri18@mhs.uinjkt.ac.id)

## Abstrak

*DOTA 2 merupakan salah satu gim yang banyak diminati baik oleh masyarakat umum maupun komunitas profesional game. Mereka saling bersaing untuk mengembangkan strategi terbaik untuk mengalahkan semua musuh yang mereka hadapi. Untuk mengembangkan strategi yang terbaik diperlukan sistem analisis yang baik dan akurat. Data mining dapat digunakan untuk memecahkan masalah ini dengan menggali informasi berharga dari dataset menggunakan metode tertentu. Metode prediksi merupakan salah satu metode dalam data mining yang paling tepat untuk mencari prediksi kemenangan untuk gim DOTA 2. Salah satu cara yang cukup sederhana dan bisa digunakan adalah Logistic Regression. Hasil penelitian ini menunjukkan bahwa Logistic Regression mampu membuat prediksi dengan baik dengan akurasi 73,7 %. Data yang digunakan dalam penelitian ini sebanyak 10.000 permainan yang diperoleh dari open data. Diharapkan penelitian ini dapat membantu pemain dalam memberikan informasi untuk mengembangkan strategi permainan dan memajukan industry eSport di Indonesia.*

**Kata kunci:** Data Mining, DOTA 2, Logistic Regression, Prediksi

## I. PENDAHULUAN

DotA 2 (Defense of the Ancients) adalah gim video dua tim, Radiant dan Dire, masing-masing dengan lima pemain, mencoba menghancurkan basis satu sama lain. Setiap pemain memilih salah satu dari 117 pahlawan dengan kemampuan dan peran magis yang berbeda. Perbandingan dapat dibuat untuk permainan catur di mana setiap bidak (seorang pemain) memiliki kemampuan serangan (peran) yang berbeda, dan bidak-bidak tersebut digunakan dalam kombinasi untuk mengeluarkan bidak lawan. 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.

Penelitian ini mencoba untuk melakukan prediksi faktor-faktor dalam kemenangan gim DOTA 2 berdasarkan data pertandingan dengan menggunakan Logistic Regression. Data pertandingan DOTA 2 didapatkan dari situs [www.kaggle.com](http://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.

## II. STUDI LITERATUR

Sebagai bahan referensi penulis menggunakan 3 jurnal :

## 1. ANALISIS FINANCIAL DISTRESS PERUSAHAAN MANUFAKTUR DI INDONESIA DENGAN REGRESI LOGISTIK

Jurnal pertama yaitu penelitian mengenai analisis financial distress perusahaan manufaktur menggunakan algoritma regresi logistik. Jurnal ini dibuat oleh Intan Zakiyatul Muflihah.

Pada bagian pembahasan didapati bahwa model mampu memprediksi financial distress perusahaan manufaktur pada periode penelitian sebesar 94,1% atau 367 dari 390 perusahaan manufaktur. Selanjutnya dapat menambahkan variabel variabel independent yang di anggap mampu memprediksi financial distress perusahaan dan juga menggunakan metode analisis statistik lainnya yang mungkin lebih akurat memprediksi financial distress perusahaan [3].

## 2. ANALISIS DIAGNOSIS PASIEN KANKER PAYUDARA MENGGUNAKAN REGRESI LOGISTIK DAN SUPPORT VECTOR MACHINE (SVM) BERDASARKAN HASIL MAMOGRAFI

Pada jurnal kedua dengan topik prediksi analisis diagnosis pasien kanker payudara dibuat oleh Fourina Ayu Novianti dan Santi Wulan Purnami. Jurnal ini juga menggunakan metode regresi logistik.

Pada bagian hasil dan pembahasan didapati bahwa metode regresi logistik dapat memprediksi diagnosis pasien kanker payudara dengan akurasi sebesar 84,90% [4].

## 3. ANALISIS REGRESI LOGISTIK UNTUK MENENTUKAN FAKTOR-FAKTOR YANG MEMPENGARUHI INDEKS PRESTASI KUMULATIF (IPK) MAHASISWA FMIPA UNIVERSITAS SAM RATULANGI MANADO

Penelitian pada jurnal ini menggunakan metode regresi logistic untuk menentukan faktor-faktor yang mempengaruhi IPK mahasiswa. Berdasarkan analisis uji parsial variabel Program studi dan tempat tinggal berpeluang memiliki

pengaruh terhadap Indeks Prestasi Kumulatif (IPK) mahasiswa di FMIPA Unsrat Manado[5].

## III. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif yang bertujuan untuk mengetahui pengaruh *Gold*, dan *XP* terhadap kemenangan. Analisis dalam penelitian ini menggunakan metode regresi logistik, dimana regresi logistik adalah salah satu metode analisis statistik yang digunakan untuk memodelkan hubungan variabel independent terhadap variabel dependent yang berskala data nominal/ordinal. Data yang digunakan dalam penelitian ini yaitu data sekunder yang berasal dari Data pertandingan DOTA 2 didapatkan dari situs [www.kaggle.com](http://www.kaggle.com) [6]. Kemudian data diagregasikan untuk setiap pemain agar mendapatkan nilai total untuk tim. Kemudian penulis mengambil sampel data sejumlah setengah dari data tersebut dan disaring agar hanya ada data dengan mode yang dapat memilih semua hero, yaitu captain mode, all pick, dan random draft.

Variabel yang digunakan dalam penelitian ini menggunakan dua variabel, yaitu variabel independen yang terdiri dari *Gold* dan *XP* sedangkan variabel dependen adalah hasil pertandingan.

## IV. HASIL DAN PEMBAHASAN

Bab ini akan menjelaskan pemrosesan data sampai mendapatkan hasil akurasi prediksi dari dataset yang digunakan menggunakan metode regresi logistik.

Ada banyak faktor-faktoe yang terjadi selama permainan selain memilih karakter, dua di antaranya adalah dapatkan gold dan *XP* dengan membunuh lawan atau monster. *Gold* dapat digunakan untuk membeli item yang meningkatkan statistik pemain, sementara *XP* menaikkan levelnya memungkinkan mereka mempelajari kemampuan yang kuat. Semakin banyak *Gold* dan *XP* yang dimiliki Radiant daripada Dire, semakin besar kemungkinannya untuk mendapatkan kemenangan.

Pemrosesan data yang akan dilakukan secara berturut-turut adalah:

### A. Persiapan Data

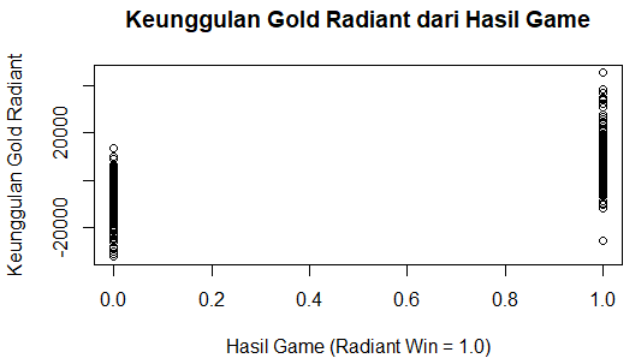
Penulis telah mengambil sampel setengah dari jumlah asli game untuk mengurangi ukuran datanya. Selain itu, penulis telah memfilter kumpulan data berisi hanya game dengan mode (*captains mode*, *all pick*, *random draft*) yang dimainkan menggunakan aturan resmi.

match_id_hash	game_time	game_mode	lobby_type	objectives_len	chat_len	r1_hero_id	r1_kills
11417	5ba547595bfa6225adcf594240fa44	126	22	7	1	4	14
5582	403c0fcd5eab3da1a103c1f1aa2179	347	22	7	1	11	4
25239	ba9fa5a9c34a7d44c54224355e032a4c	633	22	7	1	4	62
28676	ecff9569165eb7c0f8c234ef787847e0	940	22	7	3	1	87
4452	37aeeb3f6610d5a0c89ba9a74f2c2bc	1389	22	0	6	28	45
30627	b3816a237a4ab69585cfc464194da39	1683	22	0	7	0	22
34642	d9a1f6674222e5712bd72a86b99e3ff4	574	22	7	1	3	86
7516	8c896b8002874f3b2b646d0d43deec	190	22	7	1	0	15
15861	163b424e76f92ba751115a717cb422ba	2179	22	7	19	25	54
11620	1533f429519930a5f432ae37a67aa36	179	22	7	1	0	101
27130	ccdc5ef6bdc4dbbe78976edef523b992a	2814	22	7	34	7	30
30136	7add4b70cb6c39ebcbcad000d655cfc	530	22	7	2	5	16
24894	207a5be9eb394176a5492c9f1d2fdddb	196	22	0	1	1	20
31763	07ca93a242e6b347783f8284a39d6e1a	356	22	7	1	0	40
13346	c36018a9622f61c044444c8124c49d12	1708	3	7	14	0	11
12255	0643fa1aa2b56bb1a9292b099bce22d	2151	22	0	13	19	35
30925	5547a25a9eb1338ba762410bdeff9b29	939	22	7	1	0	8
9978	9e214a3149031f0099349593ca23c68	2383	2	7	26	5	76

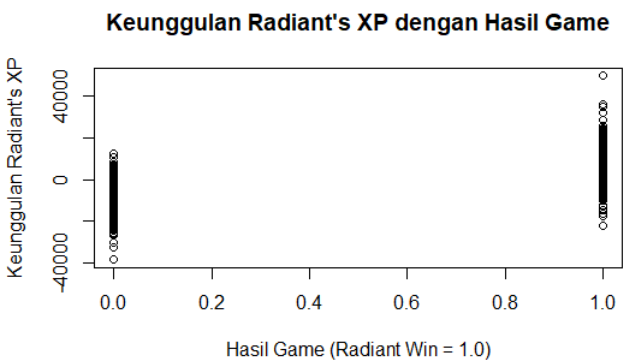
Gambar 1. Tabel Data Setelah di filter

## B. EDA

Exploratory Data Analysis (EDA) adalah bagian dari proses data science. EDA menjadi sangat penting sebelum melakukan feature engineering dan modeling karena dalam tahap ini penulis harus memahami datanya terlebih dahulu.



Gambar 2. Visualisasi data set *Gold*

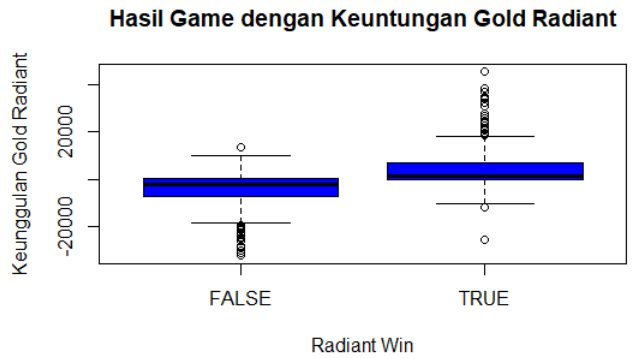


Gambar 3. Visualisasi data set *XP*

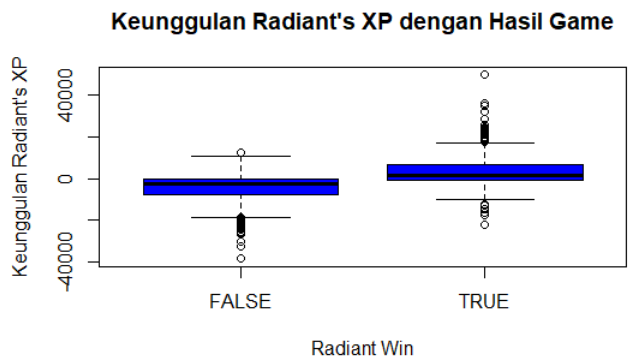


Gambar 4. Visualisasi perbedaan *Gold* dan *XP*

Dapat disimpulkan dari grafik tersebut bahwa banyak data yang saling bertumpang tindih, sehingga penulis membuat boxplot agar datanya dapat diobservasi lebih jauh.



Gambar 5. Visualisasi data set *Gold*



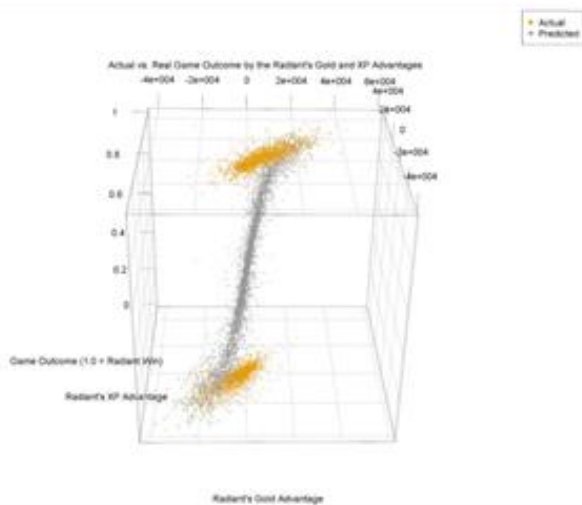
Gambar 6. Visualisasi data set *XP*

Setelah penulis memvisualisaikan dengan menggunakan boxplot, maka dapat terlihat bahwa terdapat tren pada kedua variabel yang mana setiap variabel terlihat berkorelasi positif dengan kemenangan. Walaupun banyak pencilan yang muncul, tetapi pencilan tersebut hanya 10% dari keseluruhan data.

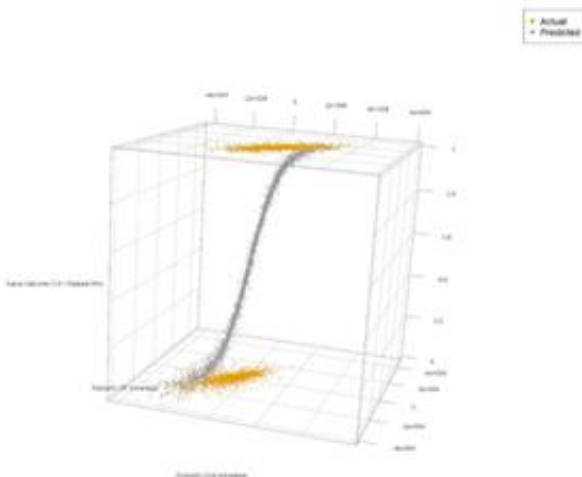
## C. Pembuatan Model

Selanjutnya penulis membuat model regresi logistik dengan variable kategorial “Menang” atau “Kalah”. Variabel independent penulis adalah

variable kontinu yaitu, *Gold* dan *XP*. Penulis membagi kumpulan data pelatihan dan pengujian dengan presentase 60%/40%. Kemudian menggunakan 3D Scatterplot untuk memvisualisasikan hasil prediksi dan hasil yang sebenarnya.



Gambar 7. Hasil Visualisasi Game yang sebenarnya vs. yang diprediksi dengan Keunggulan Gold dan XP Radiant, tampilan 1



Gambar 8. Hasil Visualisasi Game yang sebenarnya vs. yang diprediksi dengan Keunggulan Gold dan XP Radiant, tampilan 2

Perhatikan tumpang tindih antara nilai prediksi dan nilai sebenarnya. Ini selanjutnya membuktikan keakuratan file model.

## D. Hasil Perhitungan

```
Call:
glm(formula = formula, family = "binomial", data = trainGameData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7204  -0.9271   0.1449   0.9398   3.3581

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.782e-01  8.448e-02  3.293 0.000991 ***
as.numeric(total_gold_difference) 2.062e-04  2.873e-05  7.177 7.14e-13 ***
as.numeric(total_xp_difference)  7.372e-05  2.444e-05  3.017 0.002555 **

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1170.81 on 849 degrees of freedom
Residual deviance: 838.47 on 847 degrees of freedom
(9087 observations deleted due to missingness)
AIC: 844.47

Number of Fisher Scoring iterations: 6
```

Gambar 9. Hasil perhitungan regresi logistik

Intercept =  $2.782e^{-1}$

Keuntungan Gold =  $2.062e^{-4}$

Keuntungan XP =  $7.372e^{-5}$

Setelah itu, model diuji dengan data tersebut dengan batasan jika lebih dari 50% hasilnya, maka hasil prediksinya adalah menang untuk tim Radiant, jika kurang maka kalah untuk tim Radiant dan menang untuk tim Dire. Akurasi yang diraih model penulis saat tahap latih adalah 72.7%, sehingga akurasi tersebut lebih tinggi daripada peluang menebak kepala atau ekor koin (yaitu 50%:50%). Saat masuk ke tahap uji, model penulis mendapat akurasi sebesar 73.7%.

## V. KESIMPULAN DAN SARAN

Berdasarkan serangkaian kegiatan yang telah dilakukan maka dapat disimpulkan bahwa algoritma Logistic Regression dapat digunakan untuk melakukan faktor prediksi dengan baik.

Semua hipotesis terpenuhi yakni memperoleh lebih banyak gold dan XP daripada tim lain berkolerasi positif dengan kemenangan. Ini faktor penting dalam permainan seharusnya setiap karakter dapat melakukan “Last Hit” pada creep dan monster jungle. Model tersebut secara akurat memprediksi 73,7% hasil game dalam kumpulan data pengujian menggunakan klasifikasi menang / kalah 50%. Ada kemungkinan tinggi untuk menang sekitar 2000 gold dan 2000 XP poin keuntungan.

Saran kedepannya diharapkan dapat menambahkan 3 prediktor lagi seperti, *classification tree*, *pruning*, *bagging*, and *random forest* untuk modelling.

## 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] Muflihah, Intan Zakiatul. (2017). *Analisis Financial Distress Perusahaan Manufaktur Di Indonesia dengan Regresi Logistik*. Majalah Ekonomi\_ISSN No. 1411-9501\_Vol.XXII No.2. 254-269
- [4] Fourina Ayu Novianti & Santi Wulan Purnami. (2012). *Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan Support Vector Machine (SVM) Berdasarkan Hasil Mamografi*. Jurnal Sains dan Seni ITS, 1(1), 147-152.
- [5] Yumira Adriani Tampil, Hanny Komalig, & Yohanis Langi. (2017). *Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado*. jdC, 6(2), 57-63.
- [6] OpenDataScience (Diperbaharui pada Agustus 2016). *mlcourse.ai: Dota 2 Winner Prediction*. <https://www.kaggle.com/c/mlcourse-dota2-win-prediction/data>. Diakses pada tanggal 10 Desember 2020.