

## Analisis Kemenangan Pemain Pada Permainan Player Unknown Battle Grounds (PUBG) Menggunakan Metode CRISP-DM

Yehezkiel Victorious Ermanto<sup>1</sup>, Yulia Wahyuningsih<sup>2</sup>, Ronald Julio<sup>3</sup>

<sup>1</sup> Universitas Katolik Darma Cendika

<sup>2</sup> Universitas Katolik Darma Cendika

<sup>3</sup> Universitas Katolik Darma Cendika

### Abstrak

PUBG atau disebut juga dengan Player Unknown BattleGrounds merupakan game yang terdapat unsur battle royale yang bisa dimainkan melalui PC atau Smartphone. Game ini menjadi sangat populer sejak ajang e-sports yang diadakan pada akhir tahun 2000, serta banyak pemain yang berlomba-lomba mencari cara maupun strategi terbaik untuk dapat menang dalam game PUBG ini. Berdasarkan hal tersebut, kita dapat menggunakan salah satu bidang dalam ilmu komputer yaitu data mining, yang dimana memungkinkan kita untuk memprediksi tingkat kemenangan pemain dalam pertandingan PUBG. Penelitian ini mengambil dataset PUBG dari Kaggle, dalam Dataset ini memiliki banyak atribut mulai dari ID, dan sebagainya. Penelitian ini menggunakan metode CRISP-DM atau Cross Industry Standard Process for Data Mining yang dimana metode untuk melakukan pengolahan data hingga mendapatkan sebuah kesimpulan secara berkelanjutan. Tujuan dari penelitian ini adalah mencari atribut-atribut dimana beberapa atribut memiliki pengaruh yang cukup tinggi terhadap tingkat kemenangan pertandingan yaitu 'walkDistance', 'boosts', 'weaponsAcquired', 'killPlace'. ' dengan kemungkinan 90%.

**Kata kunci:** *PUBG; CRISP-DM; Correlation Matrix, Data mining*

### Abstract

**[Title: CRISP-DM Methodology to Analysis of PUBG Player's Chances of Winning]**

*PUBG or also known as Player Unknown Battle Grounds is a game that contains elements of battle royale which can be played via PC or Smartphone. This game has become very popular since the e-sports event held at the end of 2000, and many players are competing to find the best way and strategy to win in this PUBG game. Based on this, we can use one of the fields in computer science, namely data mining, which allows us to predict the win rate of players in PUBG matches. This research takes the PUBG dataset from Kaggle. In this dataset, it has many attributes ranging from ID, and so on. This study uses the CRISP-DM method or Cross Industry Standard Process for Data Mining which is a method for processing data to get a conclusion on an ongoing basis match i.e., 'walkDistance', 'boosts', 'weaponsAcquired', 'killPlace'. with 90% chance of winning the game.*

**Keywords:** *PUBG; CRISP-DM; Correlation Matrix, data mining*

## 1. PENDAHULUAN

*Player Unknown BattleGrounds* atau biasa disingkat dengan PUBG adalah game battle royale *online* yang dikembangkan oleh PUBG Corporation, game ini bisa dimainkan baik melalui *smartphone* maupun di komputer atau laptop, serta permainan ini merupakan tipe permainan FPS atau *First Person Shooter*. Pemain dapat memilih apakah mereka ingin bermain sendiri atau dengan tim yang terdiri dari 2 (dua) hingga 4 (empat) pemain. Dalam permainan ini pemain juga dapat memilih berdasarkan sudut pandangnya baik sebagai orang pertama atau FPP (*First Person Perspective*) atau sebagai sudut pandang orang ketiga atau TPP (*Third Person Perspective*) yang dimana masing-masing pilihan memiliki kelebihan dan kekurangan dalam pertempuran dan tingkat kesadaran sekelilingnya.

Pemenang dari permainan ini adalah pemain yang dapat bertahan sampai waktu permainan berakhir, serta banyak kemungkinan pemain tersebut dapat memenangkan permainan baik dari senjata yang dimilikinya atau juga teknik bermainnya seperti bersembunyi atau melakukan pengobatan

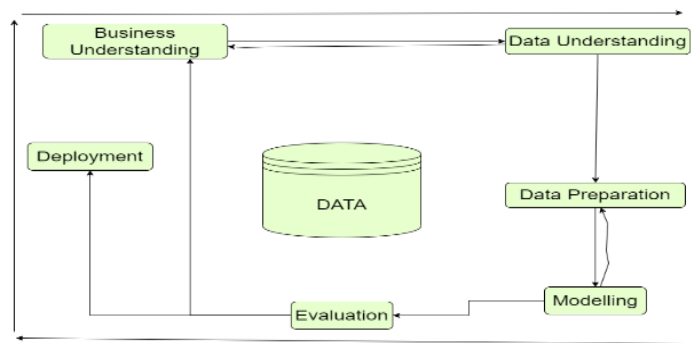
(Mamulpet, 2019). Durasi permainan PUBG ini kurang lebih 30 (tiga puluh) menit setiap pertandingan, dimulai dari pemain terjun payung dari pesawat. Kemudian ketika pemain sudah sampai di medan pertempuran, pemain harus bertahan sambil mencari perlengkapan yang diperlukan dan juga berhadapan dengan pemain lawan hingga waktu permainan selesai, dan pemain yang dapat bertahan hidup sampai waktu berakhir, maka dia menjadi pemenang. *E-sport* atau juga *Electronic Sport* merupakan salah satu jenis olahraga berbasis video game, terdapat beberapa aliran atau *genre e-sport* yang terkenal, diantaranya adalah MOBA (*Multiplayer Online Battle Arena*), FPS (*First Person Shooter*), *Fighting (Street Fighter)*, dan *Real Time Strategy Games*. *E-sport* ini juga masuk dalam salah satu cabang turnamen olahraga yang dimana turnamen *e-sport* ini berkembang pesat pada penghujung tahun 2000, kemudian pada tahun 2018 turnamen esports mencapai 380 juta penonton yang dimana sebagian besar merupakan penggemar dari *e-sport*, dan salah satu permainan yang juga hadir dalam *e-sport* adalah PUBG sebagai game baru yang cukup berpengaruh pada berkembangnya esports (Ghazali et al., 2021). Jadi dikarenakan tidak mudah untuk bertahan hidup dan menjadi pemain yang terakhir bertahan dalam permainan tersebut, perlu adanya strategi terbaik dan juga analisis dalam pencarian strategi yang baik yaitu faktor pendukung tingkat kemenangan pemain sehingga pemain dapat memperoleh kemenangan yang baik pula secara individu maupun dalam bentuk tim. Penelitian sebelumnya yang dilakukan oleh Ghazali, Sanat, dan As'ail yang mengambil topik prediksi posisi kemenangan pemain PUBG dengan *machine learning*, menghasilkan penelitian yaitu setelah melakukan investigasi dengan teknik Regression Tree untuk membuat model dengan kombinasi yang berbeda dari seleksi fitur, didapatkan atribut-atribut yang penting dalam memenangkan permainan PUBG, antara lain 'boosts', 'killPlace', 'kill', 'matchDuration', 'DBNOs', 'rideDistance', dan 'walkDistance' (Ghazali et al., 2021). Penelitian serupa mengenai prediksi permainan PUBG dilakukan oleh Mamulpet mengenai prediksi posisi pemenang permainan PUBG dengan menggunakan *Artificial Neural Network*, mendapatkan hasil penelitian yaitu terdapat beberapa atribut yang berkontribusi dalam perolehan persentil kemenangan, kemudian dengan menggunakan model *Light BGM* menghasilkan akurasi sebesar 93% dalam kumpulan data yang besar dan juga selain itu didapatkan kesimpulan bahwa pemain yang bersembunyi di zona aman selama permainan dapat juga memperoleh kemenangan.

Salah satu bidang studi dalam ilmu komputer yang dapat melakukan analisis adalah *Data Mining*. Dalam prosesnya, data mining mencari pengetahuan dari data. Data Mining dapat didefinisikan dengan berbagai cara, banyak orang menyamakan *data mining* dengan istilah populer yang digunakan yaitu KDD atau *Knowledge Discovery from Data* (Long, 2012). *Data Mining merupakan inti* dari proses KDD atau *Knowledge Discovery from Data* yang melibatkan algoritma inferring untuk mengeksplorasi data, mengembangkan model dan menemukan pola sebelumnya yang diketahui untuk membentuk sebuah model, yang dimana model tersebut digunakan untuk memahami fenomena dari data, kemudian melakukan analisis, serta melakukan prediksi (Maimon & Rokach, 2010), beberapa penelitian yang berkaitan dengan KDD yang pernah dilakukan diantaranya berdasarkan penelitian yang dilakukan oleh Nanthawadee, Choochart, Hieu Chi Dam, dan Thanaruk, yang dimana mengambil topik *multidimensional sentiment cube mining* untuk proses monitoring

untuk memonitor sentimen pelanggan dari media sosial dalam berbagai kasus bisnis. Hasil dari penelitian ini adalah dengan menggunakan MDSC mining atau *Multi-Dimensional Sentiment Cube* yang menggunakan *text mining* dan NLP, dapat mempertimbangkan *social Customer Relationship Management* dalam bisnis asuransi di Thailand, hasil MDSC juga dapat menunjukkan hubungan yang signifikan antara aspek dan sub aspek terutama dalam desain dengan analisis elemen sentimen sebagai dimensi yang dapat membantu meningkatkan sudut pandang dalam proses pemantauan (Sucharitttham et al., 2022). Penerapan CRISP-DM yang pernah dilakukan oleh Yogasetya Suhandha, dkk (Suhandha et al., 2020a) menunjukkan bahwa banyaknya mahasiswa yang lulus tidak tepat waktu disebabkan oleh IPK (Indeks Prestasi Kumulatif) yang rendah, dan adanya faktor lain. Hasil penelitian yang diperoleh berupa website yang dapat menampilkan hasil clustering siswa. Selanjutnya, penelitian yang dilakukan oleh Msy Aulia Hasanah, dkk (Hasanah et al., 2021) bertujuan untuk mengklasifikasikan hujan berdasarkan kategorinya menggunakan metode CRISP-DM dan algoritma CART (*Classification and Regression Tree*). Hasil yang diperoleh adalah akurasi sebesar 89%. Penelitian yang lain dimana juga menggunakan metode CRISP-DM, yaitu berdasarkan penelitian dari Dhewayani, dkk mengenai pengelompokan daerah rawan bencana kebakaran, peneliti ini menggunakan metode CRISP-DM dan algoritma *K-Means Clustering*, tujuan dari penelitian ini adalah untuk menghasilkan kelompok daerah yang memiliki potensi tinggi untuk terjadi kebakaran dan didapatkan hasil yaitu terdapat cluster tertingggi yaitu cluster 3 (tiga) dengan indeks kebakaran di setiap kota tercatat mencapai 87 (delapan puluh tujuh) kejadian setiap tahunnya (Dhewayani et al., n.d.). Penelitian lainnya dilakukan oleh Dwi Astuti, dkk yaitu mengenai penentuan strategi promosi UMKM yang dimana penelitian ini menggunakan metode CRISP-DM dan algoritma K-Means Clustering dengan tujuan untuk mengelompokkan UMKM yang memiliki karakteristik yang sama, sehingga dengan adanya pengelompokkan UMKM yang sesuai dapat menjadi dasar dalam pengambilan keputusan strategi promosi yang baik (Astuti et al., 2019). Berdasarkan penelitian yang dilakukan oleh Haristu, dkk mengenai prediksi rasio kemenangan pemain PUBG dengan penerapan *Random Forest*, dimana dengan menggunakan KDD sebagai proses pra pemrosesan dataset kemudian setelah ini menerapkan algoritma random forest. Hasil dari penelitian ini adalah penerapan KDD dan juga algoritma Random Forest pada penelitian ini dapat digunakan untuk melakukan prediksi rasio kemenangan pemain, dengan didapatkan akurasi sebesar 88,19% dan juga atribut yang paling berpengaruh dalam menentukan klasifikasi rasio kemenangan pemain pada permainan *Player Unknown Battlegrounds* adalah atribut 'solo\_KillDeathRatio' (Reinardus et al., n.d.). Penelitian ini dilakukan untuk menganalisis serta mencari faktor-faktor yang mempengaruhi tingkat kemenangan pemain dalam game PUBG dari berbagai sisi atau sudut pandang atau atribut-atribut yang berpengaruh, baik dari jenis peralatan yang mereka miliki maupun teknik bermain yang dilakukan oleh pemain berdasarkan dari data permainan yang telah tersedia.

## 2. METODE PENELITIAN

Berdasarkan permasalahan penelitian mengenai penerapan CRISP-DM untuk menganalisis tingkat kemenangan pemain PUBG dimulai dari bagaimana mengetahui apakah seorang pemain curang atau tidak, maka penulis mengambil suatu metode yang cocok digunakan untuk mengatasi permasalahan tersebut yaitu CRISP-DM (*Cross Industry Standard Process for Data Mining*). Metode ini menerapkan 6 tahapan dalam mengolah suatu masalah mengenai data mining, yaitu (1) Business Understanding; (2) Pemahaman Data; (3) Persiapan Data; (4) Pemodelan; (5) Evaluasi; (6) Penerapan (Suhandha et al., 2020b). Untuk penjelasan dari 6 metode tersebut akan disajikan pada gambar 1.



**Gambar 1.**

Alur Kerja CRISP-DM

1. *Business Understanding*  
Tahap ini merupakan strategi awal untuk mencapai tujuan dan menerjemahkannya ke dalam rumusan masalah *data mining*.
2. *Data Understanding*  
Tahap ini merupakan tahap pengumpulan data yang dibutuhkan dan kemudian mengevaluasi kualitas data apakah data tersebut layak digunakan atau tidak.
3. *Data Preparation*  
Tahap ini menyiapkan data yang akan digunakan dalam penelitian ini yaitu data train yang dibentuk dalam format CSV. Data ini berisi informasi mengenai permainan seperti dari ID pemain, ID pertandingan, dan lain sebagainya.
4. *Modeling*  
Pada tahap ini menggunakan masing-masing library dari Python yaitu scikit-learn, pyplot, dan sebagainya untuk melakukan pemodelan dengan menggunakan algoritma tertentu.
5. *Evaluation*  
Tahap ini mengevaluasi semua tahapan yang telah dilakukan berdasarkan tujuan semula.
6. *Deployment*  
Tahap ini menggunakan model yang telah dihasilkan. Misalnya, pembuatan laporan.

Untuk penelitian ini, penulis mengambil data dari PUBG dan menggunakan Visual Studio Code untuk menjalankan Python. Metode yang digunakan dalam penelitian ini adalah metode *Linear Regression* dan *Correlation Matrix*. Metode *linear regression* adalah metode yang digunakan untuk memprediksi suatu data apakah data tersebut dapat disimpulkan atau tidak. Metode ini juga dilakukan oleh penelitian sebelumnya yaitu penerapan *linear regression* berganda yang dilakukan oleh Muhammad Sholeh, dkk (Sholeh et al., 2022), mereka memberikan hasil bahwa nilai koefisien suatu penelitian bernilai 97%, artinya penelitian tersebut cukup optimal. Selanjutnya, penelitian yang dilakukan oleh Mirrah Zain, dkk (Zain et al., 2013) memberikan kontribusi bahwa penerapan metode regresi linier ini digunakan untuk memprediksi jumlah kelas pada periode berikutnya. Selain itu, penelitian yang dilakukan oleh Ghebyla Najla Ayuni, dkk (Ayuni & Fitriana, 2019) memprediksi penjualan properti berdasarkan data penjualan properti di PT. XYZ menggunakan metode *Linear Regression*. Hasil yang diperoleh merupakan prediksi yang diperhitungkan dengan sangat baik. Metode selanjutnya adalah

korelasi matriks. Metode selanjutnya adalah korelasi matriks. Cara ini pernah digunakan oleh Wilda Imama Sabila, dkk (Mp et al., 2022) penyebab penyakit kardiovaskuler merupakan penyakit yang cukup berbahaya sehingga hasil yang diperoleh cukup akurat.

### 3. HASIL DAN PEMBAHASAN

Dengan menggunakan metode CRISP-DM, ada beberapa proses yang dilakukan oleh penulis, diantaranya:

#### ***BUSINESS UNDERSTANDING***

Pada tahap ini merupakan bagian yang cukup penting karena membutuhkan pengetahuan dan pemahaman yang cukup terkait dengan tujuan bisnis dan tujuan pemodelan sehingga dapat membangun model terbaik, pada tahap ini tujuan dari penelitian ini adalah untuk menemukan faktor-faktor penentu kemenangan dalam permainan PUBG

#### ***DATA UNDERSTANDING***

Pada tahap ini secara garis besar kita mengecek apakah data yang dimiliki sudah baik, dimana tidak ada nilai yang kosong atau data yang tidak konsisten untuk dilanjutkan ke proses selanjutnya. Pada tahap ini, peneliti menggunakan beberapa atribut seperti pada tabel 1

**Tabel 1.**

Atribut dalam Dataset PUBG

No	Atribut	Keterangan
1	<i>Id</i>	Id pemain
2	<i>groupId</i>	Id grup
3	<i>matchId</i>	Id pertandingan
4	<i>assists</i>	Membantu rekan mengalahkan lawan
5	<i>Boosts</i>	Meningkatkan darah pemain melalui minuman energi atau peralatan lainnya seperti <i>painkillers</i>
6	<i>damageDealt</i>	Efek kerusakan dari pemain saat mengalahkan lawan
7	<i>DBNOs</i>	Jumlah dari pemain lawan yang berhasil dikalahkan namun tidak meninggal
8	<i>headshotKills</i>	Pemain mengalahkan lawan tepat di kepala
9	<i>Heals</i>	Total penyembuhan yang dilakukan pemain
10	<i>killPlace</i>	Tingkatan pemain dalam satu pertandingan
11	<i>killPoints</i>	Poin yang didapatkan ketika mengalahkan pemain lawan
12	<i>Kills</i>	Banyak dari berapa kali mengalahkan lawan
13	<i>killstreak</i>	Mengalahkan beberapa lawan dalam waktu yang singkat
14	<i>longestKill</i>	Mengalahkan lawan dengan jarak terjauh
15	<i>matchDuration</i>	Durasi dari satu pertandingan dalam satuan detik

16	<i>matchType</i>	Tipe pertandingan
17	<i>maxPlace</i>	Jumlah tempat maksimum
18	<i>numGroups</i>	Jumlah grup
19	<i>rankPoints</i>	Tingkatan poin pemain dalam permainan
20	<i>Revives</i>	Pemain menghidupkan Kembali teman timnya
21	<i>rideDistance</i>	Seberapa jauh pemain berkendara dalam satuan meter
22	<i>roadKills</i>	Pemain yang kalah oleh lawan saat berkendara
23	<i>swimDistance</i>	Seberapa jauh pemain berenang
24	<i>teamKills</i>	Jumlah dari teman yang kalah
25	<i>vehicleDestroys</i>	Jumlah dari kendaraan yang dihancurkan oleh pemain
26	<i>walkDistance</i>	Seberapa jauh pemain mengelilingi tempat pertandingan
27	<i>Acquired weapons</i>	Jumlah senjata yang dimiliki pemain
28	<i>winPoints</i>	Jumlah poin pemain yang menang
29	<i>winPlacePerc</i>	Rentang kemenangan yang diprediksi pemain

Pada dataset terdapat total 4.446.966 record dengan 29 atribut/kolom seperti terlihat pada gambar 2.

```

... Output exceeds the size limit. Open the full output data in a text editor
Id 0
groupId 0
matchId 0
assists 0
boosts 0
damageDealt 0
DBNOs 0
headshotKills 0
heals 0
killPlace 0
killPoints 0
kills 0
killStreaks 0
longestKill 0
matchDuration 0
matchType 0
maxPlace 0
numGroups 0
rankPoints 0
revives 0
rideDistance 0
roadKills 0
swimDistance 0
teamKills 0
vehicleDestroys 0
...
walkDistance 0
weaponsAcquired 0
winPoints 0
winPlacePerc 0
dtype: int64

```

**Gambar 2.**

10 Data Teratas dari Dataset

## ***DATA PREPARATION***

Pada langkah selanjutnya setelah memahami dataset dan atribut, penulis mengecek dataset yang dimiliki apakah ada missing value atau data yang tidak konsisten, jika ditemukan ada data yang hilang maka dilakukan pembersihan data, dan kolom kosong dapat diisi dengan nilai rata-rata di kolom atau dapat dihapus bagian kosong tersebut. Alasan mengapa pada tahap ini dilakukan penyiapan data antara lain pengecekan nilai kosong untuk memastikan data yang diolah bersih serta menghindari ketidakakuratan dan menjaga konsistensi data (Cazacu & Titan, 2020), langkah-langkah yang dapat dilakukan seperti pada repositori berikut ini<sup>1</sup>, kemudian untuk melakukan penghapusan nilai yang kosong dapat dilakukan seperti kode berikut ini

```
df = df.dropna(how='any', axis=0)
df.isnull().sum()
```

Kemudian setelah dijalankan tampilannya seperti gambar 3, tampak dataset yang dimiliki tidak memiliki nilai kosong atau clean sehingga dapat dilanjutkan ke tahap selanjutnya

<sup>1</sup><https://github.com/yehezkielermanto/AnalysisPUBGDataMining>

	Id	groupId	matchId	assists	boosts	damageDealt	DBNOs	headshotKills	heals	killPlace
0	7f96b2f878858a	4d4b580de459be	a10357fd1a4a91	0	0	0.000	0	0	0	60
1	eef90569b9d03c	684d5656442f9e	aeb375fc57110c	0	0	91.470	0	0	0	57
2	1eaf90ac73de72	6a4a42c3245a74	110163d8bb94ae	1	0	68.000	0	0	0	47
3	4616d365dd2853	a930a9c79cd721	f1f1f4ef412d7e	0	0	32.900	0	0	0	75
4	315c96c26c9aac	de04010b3458dd	6dc8ff871e21e6	0	0	100.000	0	0	0	45
5	ff79c12f326506	289a6836a88d27	bac52627a12114	0	0	100.000	1	1	0	44
6	95959be0e21ca3	2c485a1ad3d0f1	a8274e903927a2	0	0	0.000	0	0	0	96
7	311b84c6ff4390	eaba5fcb7fc1ae	292611730ca862	0	0	8.538	0	0	0	48
8	1a68204ccf9891	47cfbb04e1b1a2	df014fbee741c6	0	0	51.600	0	0	0	64
9	e5bb5a43587253	759bb6f7514fd2	3d3031c795305b	0	0	37.270	0	0	0	74

10 rows × 29 columns

**Gambar 3.**

Hasil Eksekusi untuk mengecek nilai yang kosong

## MODELING

Pada tahap ini, dataset yang telah disiapkan dapat menggunakan metode statistik atau machine learning. Pada tahap ini penulis melakukan pengecekan anomali dari dataset yang ada, terlebih dahulu menganalisa apakah terdapat anomali dalam mengalahkan pemain lawan. Cara yang dilakukan adalah dengan melakukan pengecekan dengan beberapa kondisi. Sebelumnya bisa divisualisasikan kolom 'kills' seperti yang terlihat pada gambar 4, dimana terlihat ada pemain yang mengalahkan musuh hingga 72 untuk nilai maksimal, dan ada juga pemain yang tidak mengalahkan musuh dengan angka tertinggi. untuk nilai minimal.



**Gambar 4.**

Visualisasi jumlah pemain yang mengalahkan lawan

Selanjutnya penulis memeriksa apakah ada pemain yang mengalahkan lebih dari 40 musuh, memiliki lebih dari 55 senjata, dan menjelajah lebih dari 100m dengan kode di bawah ini.

```
df[(df['kills'] >= 40) &
(df['weaponsAcquired'] > 55) &
(df['_totalDistance'] < 100.0)]
```

Kemudian setelah dijalankan, jika hasilnya seperti gambar 5, ditemukan data record dimana pemain mengalahkan lebih dari 40 musuh, memiliki lebih dari 55 senjata, dan menjelajah dengan jarak total lebih dari 100m sebanyak 4 pemain.

	Id	groupId	matchId	assists	boosts	damageDealt	DBNOs
156599	746aa7eabf7c86	5723e7d8250da3	f900de1ec39fa5	21	0	5479.0	0
672993	da31f191ace8ed	ce9a3c4950a8f2	17dea22cefe62a	10	0	5793.0	0
770454	2ade4369bccd12	9f9e64a3db8384	e024bf51bf1799	12	0	5557.0	0
1378200	f241fdbb4f94c6	fadbbc4cbb3e06	e024bf51bf1799	9	0	3707.0	0

4 rows x 31 columns

**Gambar 5.**

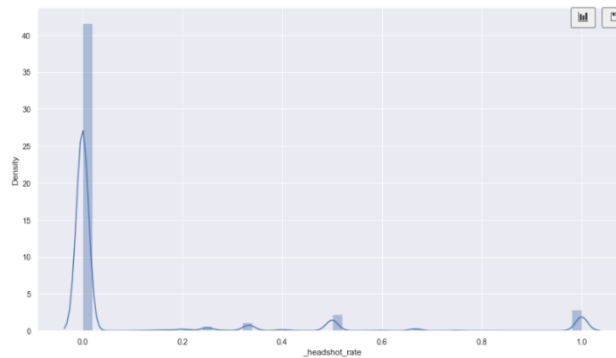
Pemain yang mengalahkan > 40, memiliki senjata > 55 dan menjelajah > 100m

Selanjutnya dilakukan pengecekan apakah ada pemain yang mengalahkan musuh lebih dari sama dengan 40 tapi tidak menggunakan *heal* atau obat-obatan dengan kode dibawah ini

```
df[(df['kills'] >= 40) & (df['heals'] == 0)]
```

Setelah menjalankan kode tersebut, muncul hasil seperti gambar 6, ditemukan lima pemain yang mengalahkan lebih dari 40 musuh tetapi tidak melakukan penyembuhan atau meminum obat-obatan.





Gambar 6.

Pemain yang tidak menggunakan *heals* dan mengalahkan > 40 pemain

Selanjutnya, anomali dari data dapat dihapus dengan kode di bawah ini

```
df.drop(df[(df['kills'] >= 40) & (df['weaponsAcquired'] > 55) &
(df['_totalDistance'] < 100.0)].index, inplace=True)

df.drop(df[(df['kills'] >= 40) & (df['heals'] == 0)].index, inplace=True)
```

Lalu yang bisa dilakukan selanjutnya adalah mengecek apakah ada pemain yang terus menerus melakukan headshot, pertama dibuat fitur baru dengan nama `_headshot_rate` yang dihitung dari kolom headshot dibagi jumlah kill musuh. Untuk visualisasi terlihat seperti gambar 7, dimana jika diamati terdapat pemain dengan headshot rate sempurna atau 1.

	Id	groupId	matchId	assists	boosts	damageDealt	DBNOs
160254	15622257cb44e2	1a513eeecfe724	db413c7c48292c	1	0	4033.0	0
2105633	770c02791306c4	48ca6706a90e10	6ee2c835176181	8	0	4272.0	40
2316123	dbc81aa64a7e3d	c50fe5c54e8bb6	5d58307bae9b50	8	0	4106.0	31
2601666	436d1530e9eb00	5c974c2bb9d9f0	fe7043ee6221c8	5	0	4347.0	0
3924729	579949f753978c	108f4c00d80882	ff9cd80c0d8fb7	1	2	3680.0	0

5 rows × 31 columns

Gambar 7.

Visualisasi pemain `_headshot_rate`

Selanjutnya penulis mengecek apakah ada pemain dengan headshot rate sempurna (1) mengalahkan lebih dari 10 musuh, dengan kode dibawah ini

```
df[(df['_headshot_rate'] == 1) & (df['kills'] >= 10)].head(10)
```

Setelah kode di eksekusi maka hasilnya seperti pada gambar 8 terdapat pemain yang memiliki `headshot_rate` 1 dan mengalahkan lebih dari 10 musuh, sebanyak 10 pemain.

	id	groupid	matchId	assists	boosts	damageDealt	DBNOs
281570	ab9d7168570927	add05ebde0214c	e016a873339c7b	2	3	1212.0	8
346124	044d18fc42fc75	fc1dbc2df6a887	628107d4c41084	3	5	1620.0	13
871244	e668a25f5488e3	5ba8feabfb2a23	f6e6581e03ba4f	0	4	1365.0	9
908815	566d8218b705aa	a9b056478d71b2	3a41552d553583	2	5	1535.0	10
963463	1bd6fd288df4f0	90584ffa22fe15	ba2de992ec7bb8	2	6	1355.0	12
1079403	1c245ed99b5f96	e42d09a9b8463a	5cec236bce68eb	0	5	1218.0	8
1167959	c4f80d4be5c561	b4a7892189b5dd	c7f7733ebbd447	0	4	1065.0	6
1348164	474a641f0a4bcb	2fdad3ca6fb3c0	114499c82f35d7	1	5	1319.0	11
1380385	202ce6a55119c5	2df66861f597b4	496700c29a5d44	1	4	1150.0	4
1483199	9d483f7cbb34d4	db5867bc814191	69495e3c478eb9	0	10	1478.0	8

10 rows × 31 columns

**Gambar 8.**

Hasil pemain memiliki `headshot_rate` 1 dan mengalahkan musuh > 10

Selanjutnya, periksa apakah ada pemain yang mengalahkan musuh tanpa bergerak terlebih dahulu. Sebelumnya penulis membuat fitur baru dengan nama `'_killsWithoutMoving'` seperti kode dibawah ini

```
df['_killsWithoutMoving'] = ((df['kills'] > 0) & (df['_totalDistance'] == 0))
```

Kemudian cek apakah ada pemain yang mengalahkan musuh tanpa bergerak dengan kode di bawah ini

```
display(df[df['_killsWithoutMoving'] == True].shape)
```

```
df[df['_killsWithoutMoving'] == True].head(10)
```

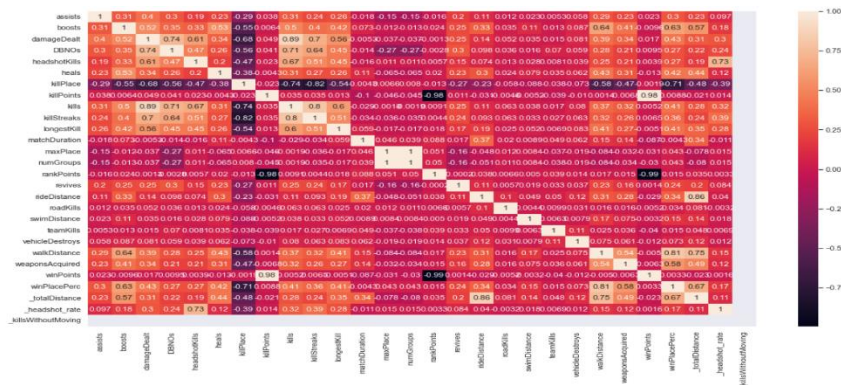
Kemudian setelah kode dijalankan maka akan muncul 10 data teratas seperti gambar 9 dari total record data 1.535

	id	groupid	matchId	assists	boosts	damageDealt	DBNOs
	1824	b538d514ef2476	0eb2ce2f43f9d6	35e7d750e442e2	0	0	593.00
	6673	6d3a61da07b7cb	2d8119b1544f87	904cecf36217df	2	0	346.60
	11892	550398a8f33db7	c3fd0e2abab0af	db6f6d1f0d4904	2	0	1750.00
	14631	58d690ee461e9d	ea5b6630b33d67	dbf34301df5e53	0	0	157.80
	15591	49b61fc963d632	0f5c5f19d9cc21	904cecf36217df	0	0	100.00
	20881	40871bf43ddac7	2cea046b7d1dce	0600f86f11c6e4	0	0	506.10
	23298	b950836d0427da	1f735b1e00d549	ad860f4e162bbc	1	0	1124.00
	24640	aeced11d46de19	d4009ffa95bb4f	73f3ed869c9171	2	0	529.90
	25659	6626c4d47cfa0	ee3fe5cd0917c3	341341834b7941	0	1	128.90
	30079	869331b90bfa3f	869ea3ad036e53	fa373e28ff5062	0	0	85.56

10 rows × 32 columns

Gambar 9.

Pemain yang mengalahkan musuh tanpa bergerak



Gambar 10.

Korelasi menggunakan Heatmap

Selanjutnya, anomali pemain mengalahkan musuh tanpa bergerak bisa dihilangkan dengan kode seperti di bawah ini

```
df.drop(df[df['_killsWithoutMoving'] == True].index,
        inplace=True)
```

Kemudian penulis juga menggunakan matriks korelasi untuk mencari hubungan yang kuat antara dua variabel, terutama dengan variabel 'winPlacePerc'. Korelasi diukur dengan statistik yang disebut koefisien korelasi, yang menggambarkan kekuatan variabel yang saling terkait, dengan rentang -1 hingga +1. Koefisien korelasi dengan nilai 0 menunjukkan bahwa tidak ada hubungan antara kedua variabel. Ada hubungan yang kuat antara variabel antara -1 dan +1. Jika koefisien korelasi adalah bilangan positif, maka variabel-variabel tersebut berhubungan langsung (misalnya jika nilai salah satu variabel meningkat, maka nilai variabel lainnya juga meningkat), sedangkan jika koefisiennya adalah bilangan negatif, maka variabel tersebut berbanding terbalik (misalnya jika nilai salah satu variabel

meningkat, nilai variabel lainnya menurun), Dalam mengukur kekuatan korelasi antara dua variabel, ada ukuran yang dapat digunakan sebagai acuan pada tabel 1 (Mukaka, 2012),

**Tabel 2.**

Representasi Ukuran Koefisien Korelasi

Ukuran Korelasi	Penafsiran
.90 hingga 1.00 (-.90 hingga -1.00)	Korelasi positif (negatif) yang sangat tinggi
.70 hingga .90 (-.70 hingga -.90)	Korelasi positif (negatif) tinggi
.50 hingga .70 (-.50 hingga -.70)	Korelasi positif (negatif) sedang
.30 hingga .50 (-.30 hingga -.50)	Korelasi positif (negatif) rendah
.00 hingga .30 (.00 hingga -.30)	Korelasi yang dapat diabaikan

Untuk visualisasi korelasi, penulis menggunakan peta panas yang merupakan metode bawaan dari perpustakaan seaborn untuk memvisualisasikan korelasi antara dua variabel yang dilambangkan dengan warna.

```
newDf = pd.read_csv('./Datasets/highly_correlated.csv') sb.set(rc = {'figure.figsize':(20,10)})
sb.heatmap(newDf.corr(),annot=True)
```

Kemudian setelah dijalankan akan terlihat seperti gambar 10, dimana tingginya korelasi warna positif yang semakin terang menunjukkan bahwa korelasi tersebut memiliki korelasi positif, sedangkan warna yang semakin gelap menunjukkan korelasi yang negatif.

Selanjutnya berdasarkan matriks korelasi yang terbentuk dengan nilai koefisien korelasi tersebut diambil variabel-variabel yang memiliki hubungan kuat dengan variabel 'winPlacePerc', antara lain variabel 'boosts', 'killPlace', 'walkDistance', dan 'weaponsAcquired'. Variabel-variabel tersebut kemudian dibandingkan dengan variabel 'winPlacePrerc' menggunakan algoritma regresi linier untuk memprediksi pemain dengan probabilitas menang yang tinggi. Berdasarkan tabel 3, pemain dengan id 'ee6a295187ba21' memiliki tingkat kemenangan yang tinggi dengan peluang 90% untuk memenangkan permainan.

**Tabel 3.**

5 Pemain dengan posisi kemenangan yang tinggi

Id	boosts	Walk Distance	Weapons	Win Place Perc

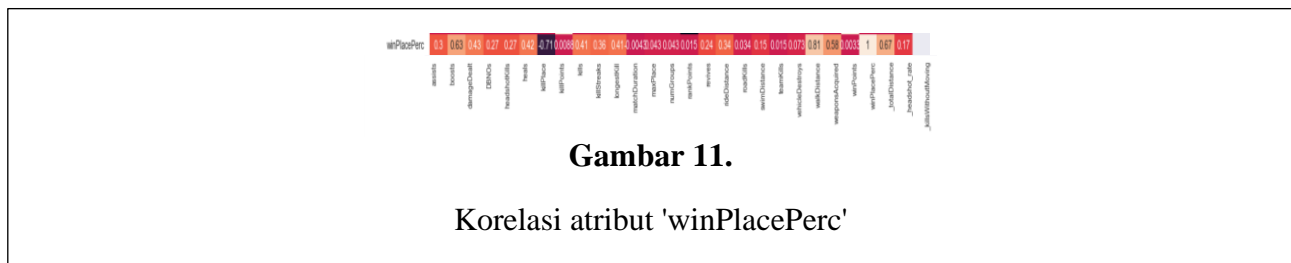
			Acquired	
9329eb41e215eb	0	588.0	1	0.294156
639bd0dcd7bda8	4	2017.0	6	0.821776
63d5c8ef8dfe91	0	787.8	4	0.475269
cf5b81422591d1	0	1812.0	3	0.609231
ee6a295187ba21	4	2963.0	4	0.902250

Kemudian dengan menggunakan *Linear Regression* didapatkan model dengan tingkat model sebesar 0.8157 untuk varians pada dataset training dan 0.8150 untuk varians pada dataset test

### EVALUATION

Pada tahap ini penulis melakukan evaluasi dimana meninjau kembali proses sebelumnya dimana matriks korelasi telah terbentuk. Evaluasi difokuskan pada atribut 'WinPlacePerc' yang memiliki korelasi dengan atribut lain seperti pada gambar 11, atribut ini dipilih karena menjadi fokus pembahasan awal dimana untuk mencari faktor-faktor yang mempengaruhi tingkat kemenangan pemain di PUBG pertandingan 'winPlacePerc' memiliki korelasi yang cukup dengan atribut 'boosts' dengan nilai korelasi positif sebesar 0,63, 'killPlace' dengan nilai korelasi negatif sebesar -0,72, 'walkDistance' dengan nilai korelasi positif 0,81, dan 'weaponAcquired' dengan nilai korelasi positif sebesar 0,58.

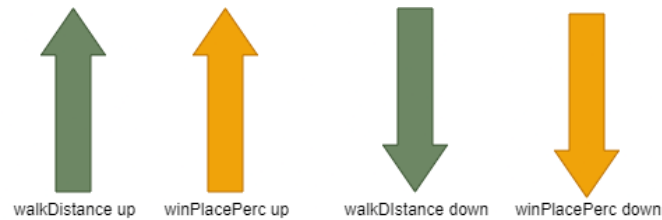
Dari korelasi yang dihasilkan semua atribut dengan atribut 'winPlacePerc', atribut 'walkDistance' merupakan atribut yang memiliki korelasi positif tinggi yang berarti jarak tempuh pemain berjalan



**Gambar 11.**

Korelasi atribut 'winPlacePerc'

kaki tinggi, dan probabilitas pemain menang posisinya juga tinggi, begitu juga sebaliknya, jika jarak tempuh pemain berjalan kaki rendah, maka kemungkinan posisi menang pemain juga rendah seperti pada gambar 12.



**Gambar 12.**

Korelasi 'walkDistance' dan 'winPlacePerc'

Kemudian jika korelasi negatif seperti antara atribut 'winPlacePerc' dan 'killPlace' dapat berarti bahwa jika peringkat pemain dalam satu permainan tinggi, kemungkinan posisi menang pemain rendah, dan jika peringkat pemain dalam permainan rendah, probabilitas posisi menang pemain tinggi seperti pada gambar 13.



**Gambar 13.**

Korelasi 'killPlace' dan 'winPlacePerc'

Pemain dengan id 'ee6a295187ba21' memiliki nilai atribut 'boosts' dengan nilai 4 yang artinya pemain ini menggunakan 4 kali boost dalam satu pertandingan, atribut 'walkDistance' dengan nilai 2.963 yang artinya pemain ini menjelajah dengan berjalan kaki sejauh 2.963 meter, dan atribut 'weaponAcquired' dengan nilai atribut 4 yang berarti pemain ini menggunakan 4 senjata dalam satu pertandingan.

#### 4. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan data pemodelan menggunakan metode CRISP-DM dan algoritma regresi linier, ditemukan beberapa determinan kemenangan pemain dalam pertandingan game PUBG dengan menggunakan matriks korelasi yang mengukur tingkat korelasi antar untuk mendapatkan beberapa atribut yang sangat berkorelasi dengan atribut 'winPlacePerc' atribut 'walkDistance' dengan nilai korelasi positif tinggi 0,81, 'boosts' dengan nilai korelasi positif 0,63, 'weaponAcquired' dengan nilai korelasi positif 0,58, dan 'killPlace' dengan negatif nilai korelasi -0,72, dan juga pemain dengan id 'ee6a295187ba21' memiliki win rate yang tinggi dengan peluang 90% untuk memenangkan permainan, selain itu kami dapat menganalisis anomali yang terdapat dalam

dataset, serta didapatkan tingkat model yang baik dengan *Linear Regression* sebesar 0.8158 varians pada dataset test.

#### DAFTAR PUSTAKA

- Astuti, D., Rahmat Iskandar, A., & Febrianti, A. (2019). *Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering*. 1(2), 60–072. <https://doi.org/10.20895/INISTA.V1I2>
- Ayuni, G. N., & Fitriana, D. (2019). Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ. *J. Telematika*, 14(2), 79–86.
- Cazacu, M., & Titan, E. (2020). Adapting CRISP-DM for Social Sciences. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 11(2sup1), 99–106. <https://doi.org/10.18662/brain/11.2Sup1/97>
- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., Jajuli, M., HSRonggo Waluyo, J., Telukjambe Timur, K., Karawang, K., & Barat, J. (n.d.). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*. <https://doi.org/10.34010/jati.v12i1>
- Ghazali<sup>1</sup>, N. F., Sanat<sup>1</sup>, N., & As', M. A. (2021). Esports Analytics on PlayerUnknown's Battlegrounds Player Placement Prediction using Machine Learning Approach. *International Journal of Human and Technology Interaction*, 5, 17–28.
- Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 5, Issue 2). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Long, T. (2012). *DATA MINING: CONCEPTS AND TECHNIQUES 3RD EDITION*.
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook (Second Edition)*.
- Mamulpet, M. M. (2019). pubg winner placement prediction using artificial neural network. in *international Journal of Engineering Applied Sciences and Technology* (Vol. 3). <http://www.ijeast.com>
- Mp, I., Harapan, M., Pasien Kardiovaskular, H., Imama, W., [57, S. ), Sabilla, W. I., Bella Vista, C., Hormansyah, D. S., Informasi, J. T., Malang, N., Soekarno, J., No, H., Lowokwaru, K., & Malang, K. (2022). Implementasi Multilayer Perceptron Untuk Memprediksi Harapan Hidup Pada Pasien Penyakit Kardiovaskular. In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 6, Issue 1).
- Mukaka, M. M. (2012). Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. In *Malawi Medical Journal* (Vol. 24, Issue 3). [www.mmj.medcol.mw](http://www.mmj.medcol.mw)
- Reinardus, ), Haristu, A., & Rosa, P. H. P. (n.d.). *Penerapan Metode Random Forest untuk Prediksi Win Ratio Pemain Player Unknown Battleground*. 4(2). [http://ejournal.ust.ac.id/index.php/Jurnal\\_Means/](http://ejournal.ust.ac.id/index.php/Jurnal_Means/)
- Sholeh, M., Rachmawati, R. Y., & Cahyo, E. N. (2022). *Penerapan Regresi Linear Ganda Untuk Memprediksi Hasil Nilai Kuesioner Mahasiswa Dengan Menggunakan Python*. 11(1), 13–24.
- Sucharittham, N., Haruechaiyasak, C., Dam, H. C., & Theeramunkong, T. (2022). Multidimensional Sentiment Cube Mining for Process Monitoring. *Trends in Sciences*, 19(9), 3682. <https://doi.org/10.48048/tis.2022.3682>

- Suhanda, Y., Kurniati, I., & Norma, S. (2020a). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika Dan Komputer*, 6(2), 12–20. <https://doi.org/10.37012/jtik.v6i2.299>
- Suhanda, Y., Kurniati, I., & Norma, S. (2020b). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika Dan Komputer*, 6(2), 12–20. <https://doi.org/10.37012/jtik.v6i2.299>
- Zain, M., Jayanti, N. K. D. A., & Atmojo, Y. P. (2013). Implementasi Forecasting Pada Perancangan Sistem Pembukaan Kelas di STIKOM Bali dengan Menggunakan Metode Regresi Linear. *Eksplora Informatika*, 3(1), 17–28.