

CHESS GAMES DATASET

Dosen Pengampu :
Agit Amrullah, S.Kom., M.Kom



Members of the group :

Henri Herdiyanto	(19.11.2934)
Mohammad Syam Caturputra	(19.11.2951)
Indra Sukma Pamungkas	(19.11.2975)
Azka Hafiyyan Muflihan	(19.11.2949)
Zainul Muttaqiin	(19.11.2973)

**S1 INFORMATIKA FAKULTAS ILMU
KOMPUTER UNIVERSITAS AMIKOM
YOGYAKARTA**

2021

A. INFORMASI DATASET

a) Informasi Umum

Dataset yang dipakai Ini adalah kumpulan lebih dari 20.000 game yang dikumpulkan dari pilihan pengguna di situs Lichess.org, dan cara mengumpulkan lebih banyak lagi. Set ini berisi:

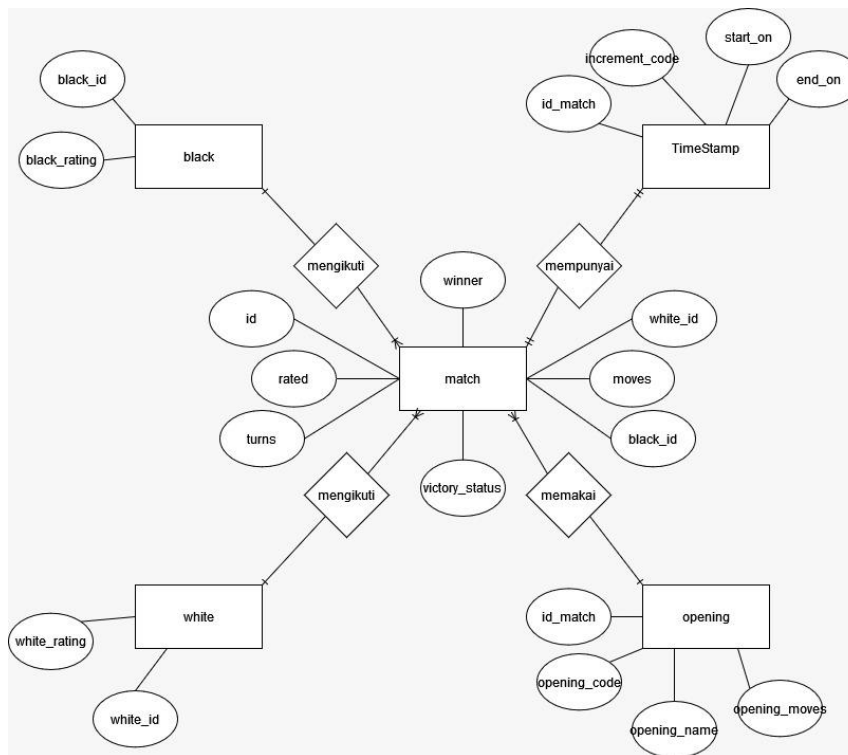
- Game ID;
- Rated (T/F);
- Start Time;
- End Time;
- Number of Turns;
- Game Status;
- Winner;
- Time Increment;
- White Player ID;
- White Player Rating;
- Black Player ID;
- Black Player Rating;
- All Moves in Standard Chess Notation;
- Opening Eco (Standardised Code for any given opening, list here);
- Opening Name;
- Opening Ply (Number of moves in the opening phase)

Untuk masing-masing game ini terpisah dari Lichess. Data ini dikumpulkan menggunakan Lichess API, yang memungkinkan pengumpulan riwayat permainan pengguna tertentu. Bagian yang sulit adalah mengumpulkan nama pengguna untuk digunakan, namun API juga memungkinkan pembuangan semua pengguna dalam tim Lichess. Ada beberapa tim di Lichess dengan lebih dari 1.500 pemain, jadi ini terbukti cara yang efektif untuk membuat pengguna mengumpulkan game.

b) Kemungkinan penggunaan

Banyak informasi terkandung dalam satu permainan catur, apalagi kumpulan data lengkap dari beberapa permainan. Ini terutama permainan pola, dan ilmu data adalah tentang mendeteksi pola dalam data, itulah sebabnya catur telah menjadi salah satu yang paling banyak diinvestasikan di bidang AI di masa lalu. Dataset ini mengumpulkan semua informasi yang tersedia dari 20.000 game dan menyajikannya dalam format yang mudah diproses untuk dianalisis, misalnya, apa yang memungkinkan pemain menang sebagai hitam atau putih, berapa banyak meta (di luar game) faktor mempengaruhi permainan, hubungan antara pembukaan dan kemenangan untuk hitam dan putih dan banyak lagi.

B. KONVERSI KE RELATIONAL DATA



Terdapat 5 Entity di dalam dataset yang saling terhubung satu sama lain dan telah di konversikan ke relational data

- | | | |
|----------|--------------|----------|
| 1. Black | 3. Timestamp | 5. Match |
| 2. White | 4. Opening | |

C. KONVERSI DATASET KE MONGODB CLUSTER

```
df = pd.read_csv('games.csv')
df.head(1)
```

	id	rated	created_at	last_move_at	turns	victory_status	winner	increment_code	white_id	white_rating	black_id	black_rating	moves	opening_eco	opening_nar
0	TZJHLJE	False	1.504210e+12	1.504210e+12	13	outoftime	white	15+2	bourgris	1500	a-00	1191	d4 d5 c4 c6 cxd5 e6 dxe6 fxe6 Nf3 Bb4+ Nc3 Ba5...	D10	Slav Defen Exchan Variati

```
[26]: df_white = pd.DataFrame(df['white_id'])
df_white.head()
```

```
[26]:
```

	white_id
0	bourgris
1	a-00
2	ischia
3	daniamurashov
4	nik221107

```
[27]: df_white["white_rating"] = df.index.map(df["white_rating"])
```

```
[28]: df["white_rating"]
```

```
[28]:
```

0	1500
1	1322
2	1496
3	1439
4	1523
...	...
20053	1691
20054	1233
20055	1219
20056	1360
20057	1235

Name: white_rating, Length: 20058, dtype: int64

```
[29]: df_white.set_index("white_id", inplace=True)
df_white
```

```
[29]:
```

	white_rating
white_id	
bourgris	1500
a-00	1322
ischia	1496
daniamurashov	1439
nik221107	1523
...	...

Database Deployments | Cloud

Project 0 Atlas App Services Charts

Database Deployments

Find a database deployment...

+ Create

Cluster0 Connect View Monitoring Browse Collections

FREE SHARED

R 9.08 W 6.7 Last 31 minutes 9.7% Connections 14.0 Last 31 minutes 14.0 In 57.5 KB/s Out 1.3 KB/s Last 31 minutes 57.5 KB/s Data Size 8.8 MB Last 31 minutes 512.0 MB Upgrade for \$9/month

- 2 GB storage
- Daily backups
- More API endpoints

Upgrade

VERSION	REGION	CLUSTER TIER	TYPE	BACKUPS	LINKED APP SERVICES	ATLAS SEARCH
5.0.9	AWS / Singapore (ap-southeast-1)	M0 Sandbox (General)	Replica Set - 3 nodes	Inactive	None Linked	Create Index

Get Started

System Status: All Good

©2022 MongoDB, Inc. Status Terms Privacy Atlas Blog Contact Sales

fp_bdl

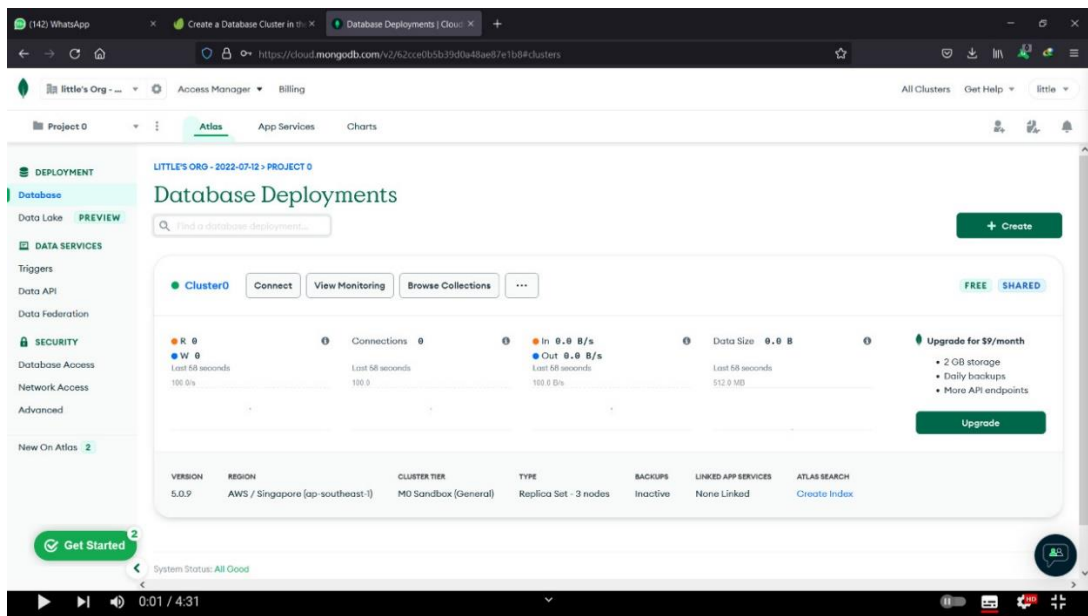
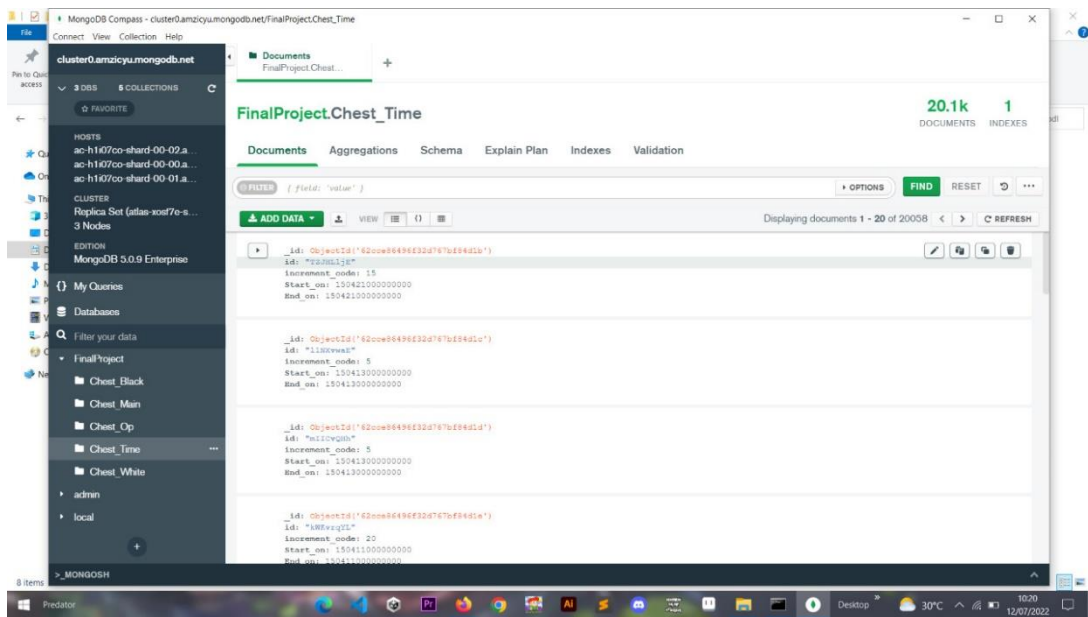
File Home Share View

File Explorer ribbon: File, Home, Share, View. Actions: Pin to Quick access, Copy, Paste, Cut, Copy path, Paste shortcut, Move to, Copy to, Delete, Rename, New folder, Easy access, Properties, Open, Select all, Select none, Invert selection.

Address bar: This PC > Documents > SEMESTER 6 > BDL > FP > fp_bdl

Name	Date modified	Type	Size
jpnyb_checkpoints	12/07/2022 9:38	File folder	
bd1_fp	12/07/2022 0:19	Jupyter Source File	61 KB
chess_clack	12/07/2022 0:17	XLS Worksheet	320 KB
chess_main	12/07/2022 0:16	XLS Worksheet	5,620 KB
chess_op	12/07/2022 0:17	XLS Worksheet	969 KB
chess_time	12/07/2022 0:18	XLS Worksheet	921 KB
chess_white	12/07/2022 0:16	XLS Worksheet	319 KB
games	11/07/2022 18:49	XLS Worksheet	7,493 KB

0 items



D. DATA MINING

a) Data Cleaning

Sebelum melakukan Representasi Data, kami terlebih dahulu melakukan data Cleaning pada data dari array of object menjadi single data yang dilakukan pada data yang di agregat kan

- col_op
- col_white
- col_black

```
[134]:
```

	_id	victory_status	winner	white_id	black_id	moves	opening
0	62cf16c7c70922607aa0212d	outoftime	white	bourgris	a-00	d4 d5 c4 c6 cxd5 e6 dxe6 fxe6 Nf3 Bb4+ Nc3 Ba5...	[{'_id': 62cce82c96f32d767bf7febf, 'id': 'TZJH...

```
[135]: opening_code = []
for i in df_match['opening']:
    c=[]
    for x in i:
        c.append(x["Opening_code"])
    opening_code.append(c)

[136]: z = []
for i in opening_code:
    z.append(i[0])
df_match['opening'] = z

[137]: df_match['opening'].head(10)

[137]: 0    D10
      1    B00
      2    C20
```

Activate Windows
Go to Settings to activate Windows.

b) Representasi Data

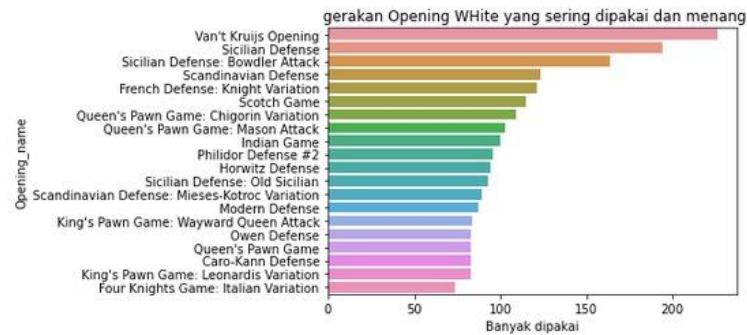
c) Table

tabel yang sudah diclean dan diambil data yang diperlukan untuk representasi

In [32]: df_match.tail()										
Out[32]:										
	_id	turns	victory_status	winner	white_id	black_id	moves	opening	white_rating	black_rating
20053	62cf16dac70922607aa06f7b	36	mate	black	mateuslichess	jamboger	e4 d6 d4 Nf6 e5 dxe5 Be2 exd4 Nf3 Nc6 O-O e5 B...	Pirc Defense	1252	1231
20054	62cf16dac70922607aa06f7c	73	mate	white	jamboger	mateuslichess	c4 d5 e3 dxc4 Bxc4 Nc6 a3 Ne5 d4 Nxc4 Nc3 e5 N...	English Opening: Anglo-Scandinavian Defense	1258	1262
20055	62cf16dac70922607aa06f65	118	mate	black	javi_r	lynnpv	e4 c5 Nh3 Nc6 Bc4 Nf6 Ng5 e6 Qf3 Ne5 Qe2 Nxc4 ...	Sicilian Defense: Brick Variation	1747	1882
20056	62cf16dac70922607aa06f6f	100	draw	draw	jamboger	nadgob	c4 e5 Nc3 Nf6 Nf3 Nc6 d4 exd4 Nxd4 Nxd4 Qxd4 c...	English Opening: King's English Variation F...	1258	1571
20057	62cf16dac70922607aa06f85	109	resign	white	marcodisogno	jamboger	e4 d6 d4 Nf6 e5 dxe5 dxe5 Qxd1+ Kxd1 Nd5 c4 Nb...	Pirc Defense	1360	1231

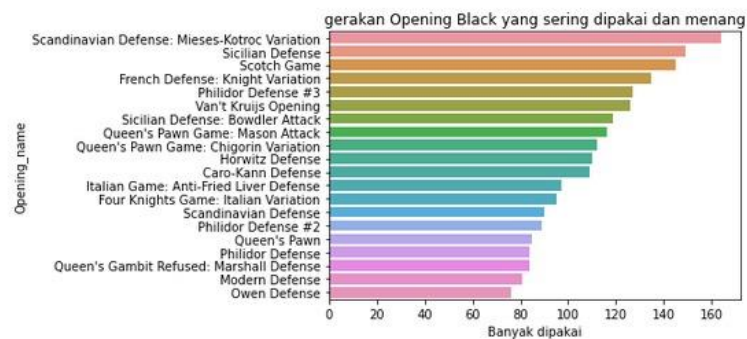
1. Gerakan White Win

```
In [49]: sns.barplot(y=black.index,x=black.values, data=df_match)
plt.title(" gerakan Opening WWhite yang sering dipakai dan menang")
plt.xlabel("Banyak dipakai")
plt.ylabel("Opening_name")
plt.show()
```



2. Gerakan Black Win

```
In [47]: sns.barplot(y=white.index,x=white.values, data=df_match)
plt.title("gerakan Opening Black yang sering dipakai dan menang")
plt.xlabel("Banyak dipakai")
plt.ylabel("Opening_name")
plt.show()
```



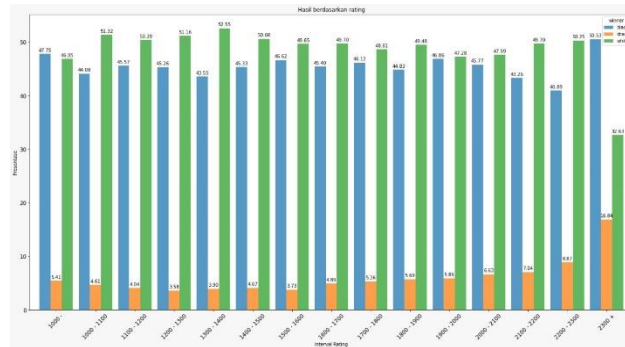
3. Persentase pertandingan berdasarkan rating

```
[188]: df_match3 = df_match2.groupby(["rating_interval", "winner"]).size().unstack().astype(int)
df_match3 = df_match3.div(df_match3.sum(axis=1), axis=0).mul(100).round(2)
```

```
[192]: plots = df_match3.plot(kind = "bar", alpha=0.75, figsize = (24, 12),
width = 0.85, title = "Hasil berdasarkan rating")

for bar in plots.patches:
    plots.annotate(format(bar.get_height(), '.2f'),
        (bar.get_x() + bar.get_width() / 2,
        bar.get_height()), ha='center', va='center',
        size=10, xytext=(0, 8),
        textcoords='offset points')
plt.xlabel("Interval Rating")
plt.ylabel("Presentase")
plt.xticks(fontsize=12, rotation = 45)
plt.yticks(fontsize=12)
plt.show()
```

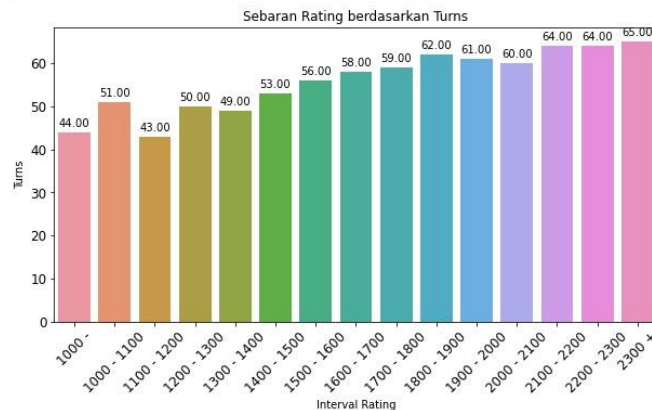
Hasil berdasarkan rating



4. Sebaran Rating Berdasarkan Turn

```
[244]: fig, ax = plt.subplots(figsize=(10,5))
plots = sns.barplot(x = df_match2.groupby("rating_interval")["turns"].mean().index,
y=df_match2.groupby("rating_interval")["turns"].median())

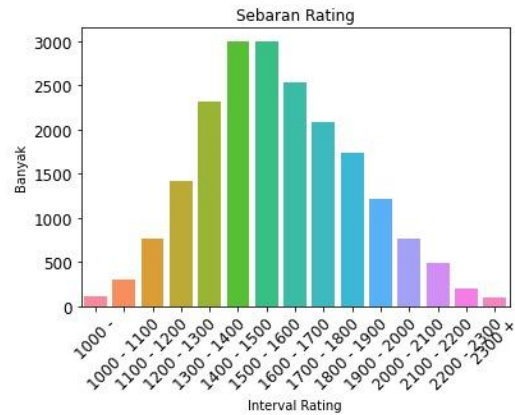
for bar in plots.patches:
    plots.annotate(format(bar.get_height(), '.2f'),
        (bar.get_x() + bar.get_width() / 2,
        bar.get_height()), ha='center', va='center',
        size=10, xytext=(0, 8),
        textcoords='offset points')
plt.title("Sebaran Rating berdasarkan Turns")
plt.xlabel("Interval Rating")
plt.ylabel("Turns")
plt.xticks(fontsize=12, rotation = 45)
plt.yticks(fontsize=12)
plt.show()
```



5. Sebaran Rating Pada User

```
sns.barplot(x = df_match2["rating_interval"].value_counts().index, y=df_match2["rating_interval"].value_counts().values,
            linewidth=1.5, saturation=1.5, order = sorted(set(df_match2["rating_interval"])))

plt.title("Sebaran Rating")
plt.xlabel("Interval Rating")
plt.ylabel("Banyak")
plt.xticks(fontsize=12, rotation = 45)
plt.yticks(fontsize=12)
plt.show()
```



E. KESIMPULAN

1. Rating User paling banyak di rentang 1400 Sampai 1600
2. Opening play paling sering menang untuk putih adalah Vant's Kruijs Opening
3. Opening play paling sering menang untuk hitam adalah Scandinavian Defense:
Mieses-Kotroc Variation
4. Berdasarkan data rata2 persentase kemenangan adalah putih
5. Semakin tinggi rating semakin besar kemungkinan draw dan juga semakin banyak giliran(turns) ada