

Nama : Muhammad Hisyam Kamil
NIM : 202210370311060
Mata Kuliah : Data, Informasi, dan Pengetahuan
Kelas : 6B
Source Code : <https://github.com/hisyam99/protondb-scraper-etl>

Tugas 1

Laporan Tugas ETL (Extract, Transform, Load) Data, Informasi, dan Pengetahuan

Analisis Data Ulasan ProtonDB (per 2 Maret 2025)

Laporan ini menyajikan analisis sangat mendalam terhadap sekitar kurang lebih 4.000+ ulasan pengguna dari ProtonDB, sebuah platform yang melacak kompatibilitas dan kinerja game yang dijalankan di Linux menggunakan lapisan kompatibilitas Proton, terutama untuk Steam Deck dan sistem berbasis Linux lainnya. Analisis ini dilakukan menggunakan proses *Extract, Transform, Load* (ETL) yang diimplementasikan dalam Python, memanfaatkan alat pemrosesan bahasa alami (NLP) dari NLTK, analisis sentimen dengan VADER, dan *library* visualisasi seperti Matplotlib dan Seaborn. Dataset, yang dibatasi hingga maksimal 500 game, memberikan wawasan tentang masalah yang dilaporkan pengguna, sentimen, istilah terkait kinerja, dan pola linguistik di seluruh ulasan yang dikategorikan berdasarkan topik seperti kompatibilitas, bug, dan kinerja.

Temuan utama mengungkapkan sentimen yang didominasi positif (48,9%) terhadap kompatibilitas Proton, dengan "layar hitam" (*black screen*) dan "*game crash*" sebagai masalah teknis yang paling sering dilaporkan. Laporan ini menawarkan rincian data yang menyeluruh, termasuk visualisasi dan wawasan statistik, untuk memandu pengembang, *gamer*, dan pengelola platform dalam memahami pengalaman pengguna dan meningkatkan fungsionalitas Proton. Laporan ini menggabungkan informasi dari dua sumber: analisis asli yang dilakukan dengan kode Python yang disediakan dan ringkasan analisis yang lebih ringkas dalam Bahasa Inggris.

1. Pendahuluan

ProtonDB adalah sumber daya yang sangat penting bagi para *gamer* yang menggunakan sistem berbasis Linux, khususnya Steam Deck. Platform ini menyediakan laporan berbasis komunitas tentang kompatibilitas dan kinerja game di bawah lapisan kompatibilitas Proton. Lapisan ini memungkinkan game Windows untuk dijalankan di Linux melalui Steam, menjembatani kesenjangan bagi pengguna di ekosistem *gaming* Linux. Hingga 2 Maret 2025, analisis ini berfokus pada ekstraksi, transformasi, dan pemuatan data dari ProtonDB untuk mengungkap pola, sentimen, dan tantangan teknis yang dilaporkan oleh pengguna di sekitar 4.612 ulasan dari hingga 500 game.

Proses ETL melibatkan pengambilan data game dan laporan melalui panggilan API, mentransformasikan data teks menggunakan teknik NLP (tokenisasi, *stemming*, *lemmatization*,

analisis sentimen, dan kategorisasi topik), dan memuat hasilnya ke dalam format terstruktur (CSV, *database* SQLite, dan visualisasi). Analisis mengkategorikan masalah ke dalam kinerja, bug, kompatibilitas, dan kombinasinya, serta memeriksa sentimen (positif, negatif, netral) dan fitur linguistik (misalnya, jenis kata, frekuensi kata).

Laporan ini bertujuan untuk memberikan tinjauan yang sangat komprehensif tentang temuan, menawarkan wawasan yang dapat ditindaklanjuti untuk pengembang, *gamer*, dan pengelola platform untuk meningkatkan kompatibilitas Proton dan mengatasi masalah teknis yang umum. Laporan ini jauh lebih rinci dan mendalam daripada ringkasan eksekutif, menggali setiap aspek analisis secara menyeluruh.

2. Metodologi

2.1 Ekstraksi Data

Data diekstraksi dari API ProtonDB (<https://protondb.max-p.me/games/> dan *endpoint* terkait) menggunakan *library requests* Python. Prosesnya mengambil:

- * Daftar game (hingga 500, dibatasi karena keterbatasan komputasi). Pembatasan ini penting untuk memastikan analisis dapat diselesaikan dalam waktu yang wajar dengan sumber daya yang tersedia.
- * Laporan individual untuk setiap game, yang berisi catatan pengguna yang menjelaskan pengalaman dengan kompatibilitas game, bug, dan kinerja. Catatan ini adalah sumber utama informasi kualitatif dan kuantitatif untuk analisis.

Data yang diekstraksi disimpan dalam file JSON untuk persistensi dan penggunaan selanjutnya, memastikan bahwa transformasi berikutnya dapat memanfaatkan data yang di-*cache* jika tersedia. Pendekatan ini mengoptimalkan proses ETL dengan menghindari pengambilan data berulang yang tidak perlu.

2.2 Transformasi Data

Fase transformasi menggunakan NLTK untuk tugas NLP tingkat lanjut, termasuk:

- * **Tokenisasi:** Memecah teks menjadi kata-kata dan kalimat individual. Ini adalah langkah fundamental dalam memproses data teks, memungkinkan analisis pada tingkat granularitas yang lebih halus.
- * **Stemming dan Lemmatization:** Mengubah kata-kata menjadi bentuk dasarnya (misalnya, "running" → "run"). *Stemming* menggunakan algoritma heuristik, sedangkan *lemmatization* menggunakan kamus dan analisis morfologis. *Lemmatization* umumnya lebih akurat tetapi lebih lambat. Kedua teknik ini membantu mengurangi variasi kata dan meningkatkan akurasi analisis frekuensi.
- * **Part-of-Speech (POS) Tagging:** Mengidentifikasi kata benda, kata kerja, kata sifat, dan kata keterangan. Ini memungkinkan analisis pola penggunaan bahasa berdasarkan jenis kata, yang dapat memberikan wawasan tentang bagaimana pengguna mengekspresikan sentimen dan menggambarkan masalah.

- * **Analisis Sentimen:** Menggunakan VADER (Valence Aware Dictionary and sEntiment Reasoner) untuk menetapkan skor sentimen (positif, negatif, netral) berdasarkan skor gabungan. Skor gabungan adalah metrik tunggal yang mewakili sentimen keseluruhan dari sebuah teks. Ambang batas berikut digunakan:
 - Positif: skor gabungan > 0.05
 - Negatif: skor gabungan < -0.05
 - Netral: skor gabungan antara -0.05 dan 0.05
- * **Kategorisasi Topik:** Mengklasifikasikan ulasan ke dalam kategori (kinerja, bug, kompatibilitas, atau kombinasinya) menggunakan set kata kunci yang telah ditentukan. Contoh kata kunci:
 - **Kinerja:** "lag", "slow", "fast", "smooth", "fps", "stutter"
 - **Bug:** "crash", "bug", "error", "broken", "fail", "glitch"
 - **Kompatibilitas:** "works", "support", "compatible", "install", "runs", "launch"
 - Jika tidak ada kata kunci yang cocok, ulasan dikategorikan sebagai "Lainnya".

Metrik tambahan yang dihitung meliputi:

- * **Jumlah Kata:** Ukuran panjang ulasan.
- * **Jumlah Karakter:** Ukuran lain dari panjang ulasan, yang mencakup spasi dan tanda baca.
- * **Jumlah Kalimat:** Jumlah kalimat dalam ulasan.
- * **Panjang Kata Rata-rata:** Rata-rata jumlah karakter per kata.
- * **Keanekaragaman Leksikal:** Rasio kata unik terhadap jumlah total kata. Nilai yang lebih tinggi menunjukkan variasi kosakata yang lebih besar.
- * **Named Entity Recognition (NER):** Mengidentifikasi entitas bernama (misalnya, nama game, perusahaan, teknologi) dalam teks.

Frekuensi kata, *bigram* (dua kata berurutan), dan *trigram* (tiga kata berurutan) juga dihitung untuk mengidentifikasi pola umum dan masalah teknis yang sering dilaporkan.

2.3 Pemuatan dan Visualisasi Data

Data yang telah ditransformasi dimuat ke dalam:

- * **File CSV:** Untuk analisis tabel yang mudah dan *import* ke *spreadsheet* atau alat analisis lainnya.

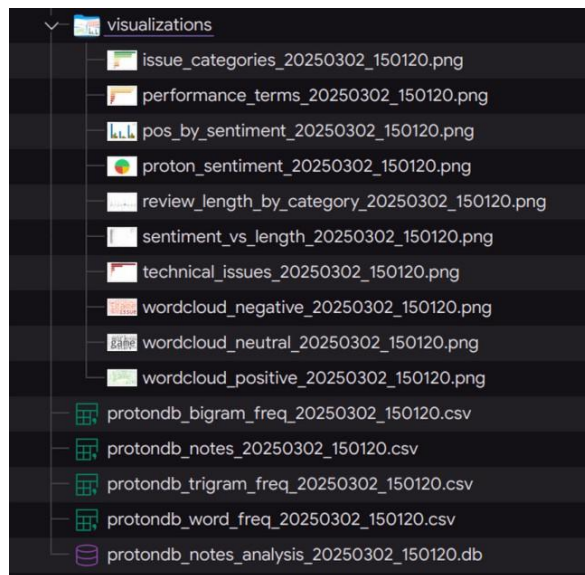
	A	B	C	D	E	F	G	H	I	J	K
1	note_text	word_count	char_count	sentence_count	avg_word_length	lexical_diversity	tokens	stemmed_tokens	lemmatized_tokens	noun_count	verb_count
2	runs perfectly without	4	31	1	7	1	runs perfectly without	run perfectly without	run perfectly without	1	
3	The same goes for s										
4	The Motivation for g	45	610	2	6.666666667	0.866666667	singleplayer works f	singleplay work fine	singleplayer work fine	20	
5	No issues, but no d3c	5	61	1	6.8	1	issues software mode	issu softwar mode av	issue software mode	3	
6	What works - Installation - Sound										
7	What does not - There is no video - The game is unplay										
8	Just like the WineHQ										
9	(To install, simply go	20	309	1	5.55	0.95	works installation sou	work instal sound vid	work installation sour	10	
10	won't start. steam shc	11	104	2	4.636363636	1	wo start steam shows	wo start steam show	wo start steam show	4	
11	Starts but the window	9	85	2	5.666666667	1	starts window becom	start window becom	start window becom	4	
12	Game works native o	16	180	3	6.375	0.875	game works native lii	game work nativ linu	game work native lin	5	
13	Game works native o	14	156	3	5.857142857	0.9285714286	game works native lii	game work nativ linu	game work native lin	4	
14	Game works native o	14	152	3	5.857142857	0.9285714286	game works native lii	game work nativ linu	game work native lin	4	
15	Game works native o	14	152	3	5.857142857	0.9285714286	game works native lii	game work nativ linu	game work native lin	4	
16	Game works native o	16	180	3	6.375	0.875	game works native lii	game work nativ linu	game work native lin	5	
17	Game works native o	23	292	3	6.565217391	0.8695652174	game works native lii	game work nativ linu	game work native lin	10	
18	Game works native o	14	152	3	5.857142857	0.9285714286	game works native lii	game work nativ linu	game work native lin	4	
19	Game works native o	23	292	3	6.565217391	0.8695652174	game works native lii	game work nativ linu	game work native lin	10	
20	Game works native o	14	152	3	5.857142857	0.9285714286	game works native lii	game work nativ linu	game work native lin	4	
21	Game works native o	23	292	3	6.565217391	0.8695652174	game works native lii	game work nativ linu	game work native lin	10	
22	Game works native o	14	152	3	5.857142857	0.9285714286	game works native lii	game work nativ linu	game work native lin	4	
23	Runs excellent	2	14	1		1	runs excellent	run excel	run excellent	2	
24	Game works native o	23	292	3	6.565217391	0.8695652174	game works native lii	game work nativ linu	game work native lin	10	

- * **Database SQLite:** Untuk kueri terstruktur yang lebih kompleks dan pengelolaan data yang efisien.

	A	B	C	D	E	F	G	H	I	J	K
1	note_text	word_count	char_count	sentence_count	avg_word_length	lexical_diversity	tokens	stemmed_tokens	lemmatized_tokens	noun_count	verb_count
2	runs perfectly without	4	31	1	7	1	runs perfectly without	run perfectly without	run perfectly without	1	
3	The same goes for s										
4	The Motivation for g	45	610	2	6.666666667	0.866666667	singleplayer works f	singleplay work fine	singleplayer work fine	20	
5	No issues, but no d3c	5	61	1	6.8	1	issues software mode	issu softwar mode av	issue software mode	3	
6	What works - Installation - Sound										
7	What does not - There is no video - The game is unplay										
8	Just like the WineHQ										
9	(To install, simply go	20	309	1	5.55	0.95	works installation sou	work instal sound vid	work installation sour	10	
10	won't start. steam shc	11	104	2	4.636363636	1	wo start steam shows	wo start steam show	wo start steam show	4	
11	Starts but the window	9	85	2	5.666666667	1	starts window becom	start window becom	start window becom	4	
12	Game works native o	16	180	3	6.375	0.875	game works native lii	game work nativ linu	game work native lin	5	
13	Game works native o	14	156	3	5.857142857	0.9285714286	game works native lii	game work nativ linu	game work native lin	4	
14	Game works native o	14	152	3	5.857142857	0.9285714286	game works native lii	game work nativ linu	game work native lin	4	
15	Game works native o	14	152	3	5.857142857	0.9285714286	game works native lii	game work nativ linu	game work native lin	4	
16	Game works native o	16	180	3	6.375	0.875	game works native lii	game work nativ linu	game work native lin	5	
17	Game works native o	23	292	3	6.565217391	0.8695652174	game works native lii	game work nativ linu	game work native lin	10	
18	Game works native o	14	152	3	5.857142857	0.9285714286	game works native lii	game work nativ linu	game work native lin	4	
19	Game works native o	23	292	3	6.565217391	0.8695652174	game works native lii	game work nativ linu	game work native lin	10	
20	Game works native o	14	152	3	5.857142857	0.9285714286	game works native lii	game work nativ linu	game work native lin	4	
21	Game works native o	23	292	3	6.565217391	0.8695652174	game works native lii	game work nativ linu	game work native lin	10	
22	Game works native o	14	152	3	5.857142857	0.9285714286	game works native lii	game work nativ linu	game work native lin	4	
23	Runs excellent	2	14	1		1	runs excellent	run excel	run excellent	2	
24	Game works native o	23	292	3	6.565217391	0.8695652174	game works native lii	game work nativ linu	game work native lin	10	

- * **Visualisasi:** Dibuat dengan Matplotlib dan Seaborn, termasuk:

- o Diagram batang
- o Diagram lingkaran
- o *Box plot*
- o *Scatter plot*
- o *Word cloud*



2.4 Alur Kerja

- * **Eksekusi `main.py`:** Pengguna menjalankan skrip `main.py` dari *command line* atau *terminal*.
- * **Menu Interaktif:** `main.py` menampilkan menu interaktif yang memungkinkan pengguna memilih salah satu dari tiga opsi:

```

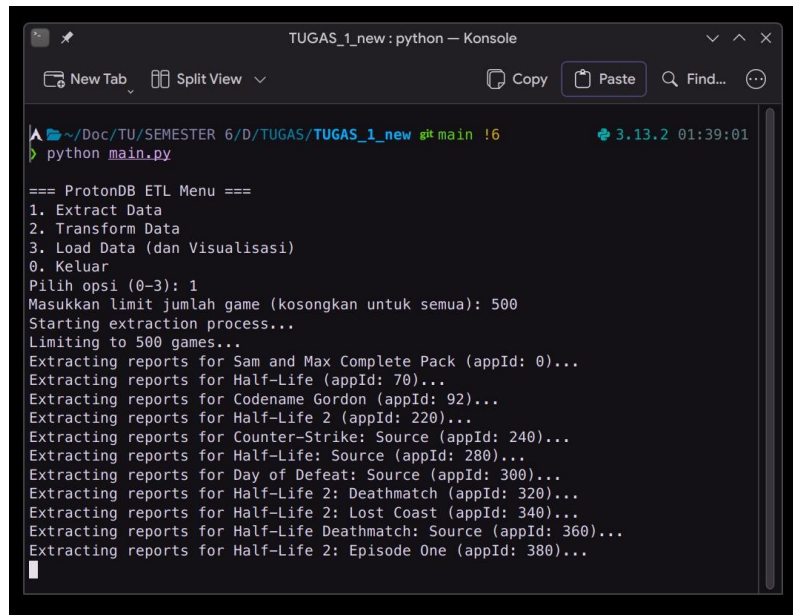
TUGAS_1_new : python - Konsole
New Tab Split View Copy Paste Find...
~/Doc/TU/SEMESTER 6/D/TUGAS/TUGAS_1_new git main !6 ?2 3.13.2 02:21:37
> python main.py

=== ProtonDB ETL Menu ===
1. Extract Data
2. Transform Data
3. Load Data (dan Visualisasi)
0. Keluar
Pilih opsi (0-3): 

```

- **1. Extract Data:** Memulai proses ekstraksi data dari API ProtonDB.
- **2. Transform Data:** Memproses data yang telah diekstraksi (atau dimuat dari file sementara) menggunakan teknik NLP.
- **3. Load Data (dan Visualisasi):** Menyimpan data yang telah ditransformasi ke dalam file CSV dan database SQLite, serta menghasilkan visualisasi.
- **0. Keluar:** Menghentikan program.

★ **Ekstraksi Data (`extract.py`):**

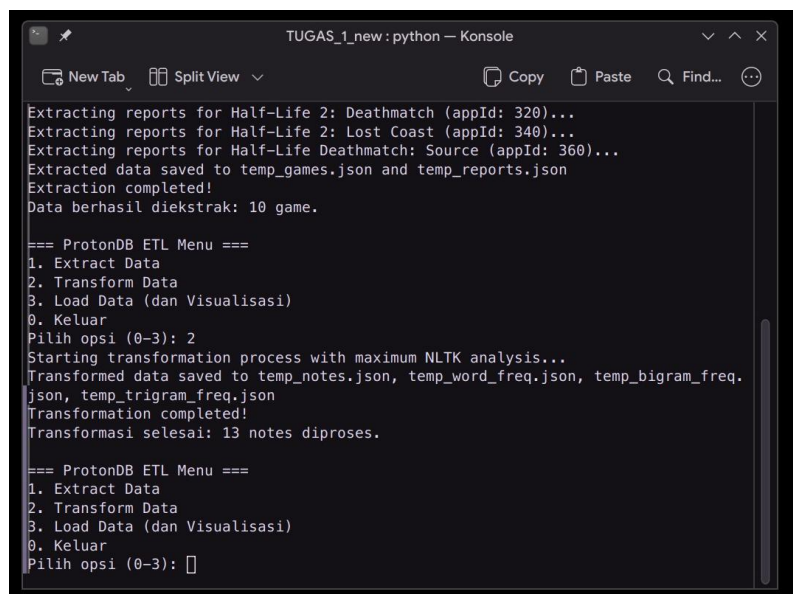


```
TUGAS_1_new: python — Konsole
New Tab Split View Copy Paste Find...
~/Doc/TU/SEMESTER 6/D/TUGAS/TUGAS_1_new git main !6 3.13.2 01:39:01
> python main.py

=== ProtonDB ETL Menu ===
1. Extract Data
2. Transform Data
3. Load Data (dan Visualisasi)
0. Keluar
Pilih opsi (0-3): 1
Masukkan limit jumlah game (kosongkan untuk semua): 500
Starting extraction process...
Limiting to 500 games...
Extracting reports for Sam and Max Complete Pack (appId: 0)...
Extracting reports for Half-Life (appId: 70)...
Extracting reports for Codename Gordon (appId: 92)...
Extracting reports for Half-Life 2 (appId: 220)...
Extracting reports for Counter-Strike: Source (appId: 240)...
Extracting reports for Half-Life: Source (appId: 280)...
Extracting reports for Day of Defeat: Source (appId: 300)...
Extracting reports for Half-Life 2: Deathmatch (appId: 320)...
Extracting reports for Half-Life 2: Lost Coast (appId: 340)...
Extracting reports for Half-Life Deathmatch: Source (appId: 360)...
Extracting reports for Half-Life 2: Episode One (appId: 380)...
```

- Jika pengguna memilih opsi 1, fungsi `extract()` dalam `extract.py` dipanggil.
- Fungsi ini pertama-tama memeriksa apakah file `temp_games.json` dan `temp_reports.json` sudah ada. Jika ya, data dimuat dari file-file ini, melewati langkah ekstraksi.
- Jika tidak, fungsi `extract_games()` dan `extract_reports()` dipanggil untuk mengambil data dari API ProtonDB menggunakan *library requests*.
- Data yang diekstraksi disimpan dalam file JSON sementara (`temp_games.json` dan `temp_reports.json`) untuk digunakan nanti.

★ **Transformasi Data (`transform.py`):**



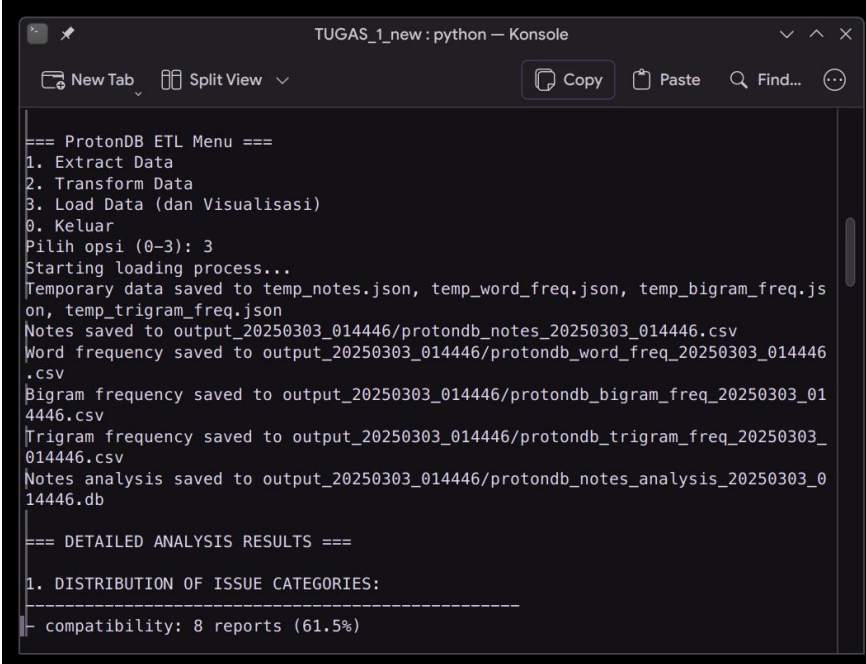
```
TUGAS_1_new: python — Konsole
New Tab Split View Copy Paste Find...
Extracting reports for Half-Life 2: Deathmatch (appId: 320)...
Extracting reports for Half-Life 2: Lost Coast (appId: 340)...
Extracting reports for Half-Life Deathmatch: Source (appId: 360)...
Extracted data saved to temp_games.json and temp_reports.json
Extraction completed!
Data berhasil diekstrak: 10 game.

=== ProtonDB ETL Menu ===
1. Extract Data
2. Transform Data
3. Load Data (dan Visualisasi)
0. Keluar
Pilih opsi (0-3): 2
Starting transformation process with maximum NLTK analysis...
Transformed data saved to temp_notes.json, temp_word_freq.json, temp_bigram_freq.
json, temp_trigram_freq.json
Transformation completed!
Transformasi selesai: 13 notes diproses.

=== ProtonDB ETL Menu ===
1. Extract Data
2. Transform Data
3. Load Data (dan Visualisasi)
0. Keluar
Pilih opsi (0-3):
```

- Jika pengguna memilih opsi 2, fungsi `transform()` dalam `transform.py` dipanggil.
- Fungsi ini juga memeriksa apakah *file-file* sementara hasil transformasi (`temp_notes.json`, `temp_word_freq.json`, `temp_bigram_freq.json`, `temp_trigram_freq.json`) sudah ada. Jika ya, data dimuat dari *file-file* ini.
- Jika tidak, fungsi ini melakukan serangkaian operasi NLP pada data yang diekstraksi, termasuk tokenisasi, *stemming*, *lemmatization*, *POS tagging*, NER, analisis sentimen, dan kategorisasi topik.
- Hasil transformasi disimpan dalam file JSON sementara.

★ **Pemuatan dan Visualisasi Data (`load.py`):**



```

TUGAS_1_new : python — Konsole

New Tab Split View Copy Paste Find...

=== ProtonDB ETL Menu ===
1. Extract Data
2. Transform Data
3. Load Data (dan Visualisasi)
0. Keluar
Pilih opsi (0-3): 3
Starting loading process...
Temporary data saved to temp_notes.json, temp_word_freq.json, temp_bigram_freq.json, temp_trigram_freq.json
Notes saved to output_20250303_014446/protondb_notes_20250303_014446.csv
Word frequency saved to output_20250303_014446/protondb_word_freq_20250303_014446.csv
Bigram frequency saved to output_20250303_014446/protondb_bigram_freq_20250303_014446.csv
Trigram frequency saved to output_20250303_014446/protondb_trigram_freq_20250303_014446.csv
Notes analysis saved to output_20250303_014446/protondb_notes_analysis_20250303_014446.db

=== DETAILED ANALYSIS RESULTS ===
1. DISTRIBUTION OF ISSUE CATEGORIES:
-----
- compatibility: 8 reports (61.5%)

```

- Jika pengguna memilih opsi 3, fungsi `load()` dalam `load.py` dipanggil.
 - Fungsi ini memuat data yang telah ditransformasi dari file JSON sementara.
 - Data tersebut kemudian disimpan secara permanen dalam file CSV dan database SQLite.
 - Berbagai visualisasi (diagram batang, diagram lingkaran, *box plot*, *scatter plot*, *word cloud*) dihasilkan menggunakan library `matplotlib` dan `seaborn`.
 - Visualisasi disimpan dalam direktori `visualizations` di dalam direktori output yang dibuat berdasarkan *timestamp*.
- ★ **Loop Menu:** Setelah setiap operasi selesai, menu ditampilkan lagi, memungkinkan pengguna untuk melakukan tindakan lain atau keluar dari program.

3. Hasil Analisis Terperinci

3.1 Distribusi Kategori Masalah

Analisis mengkategorikan masalah yang dilaporkan pengguna ke dalam delapan kategori berbeda berdasarkan keberadaan kata kunci yang terkait dengan kinerja, bug, dan kompatibilitas. Distribusi laporan adalah sebagai berikut:

Kategori Masalah	Jumlah Laporan	Persentase
Kompatibilitas	1273	27.6%
Lainnya	1047	22.7%
Kinerja, Kompatibilitas	1017	22.1%
Bug, Kompatibilitas	356	7.7%
Bug	310	6.7%
Kinerja, Bug, Kompatibilitas	287	6.2%
Kinerja	281	6.1%
Kinerja, Bug	41	0.9%

- * **Kompatibilitas (27.6%):** Kategori yang paling umum, menunjukkan bahwa pengguna sering mendiskusikan kemampuan Proton untuk menjalankan game di Linux. Ini menyoroti peran utama Proton sebagai lapisan kompatibilitas.
- * **Lainnya (22.7%):** Kategori *catch-all* untuk ulasan yang tidak sesuai dengan topik yang telah ditentukan, menunjukkan masalah yang beragam atau ambigu. Analisis lebih lanjut dari ulasan ini mungkin diperlukan untuk memahami sifat masalah yang dilaporkan.
- * **Kinerja, Kompatibilitas (22.1%):** Menggabungkan masalah kinerja dan kompatibilitas, seringkali terkait dengan seberapa baik game berjalan di samping tantangan kompatibilitas. Ini menunjukkan hubungan yang erat antara kinerja dan kompatibilitas.
- * **Bug, Kompatibilitas (7.7%):** Masalah yang melibatkan bug (misalnya, *crash*, *glitch*) di samping masalah kompatibilitas.
- * **Bug (6.7%):** Laporan bug yang berdiri sendiri, seperti *crash* atau *freeze*, tanpa penekanan kompatibilitas.
- * **Kinerja, Bug, Kompatibilitas (6.2%):** Kategori kompleks yang menggabungkan ketiga jenis masalah, menunjukkan masalah multifaset.
- * **Kinerja (6.1%):** Berfokus pada masalah kinerja seperti *lag* atau penurunan *frame rate*.
- * **Kinerja, Bug (0.9%):** Kategori yang paling jarang, membahas kinerja dan bug tanpa masalah kompatibilitas.

Distribusi ini, divisualisasikan dalam diagram batang ("Distribusi Kategori Masalah di ProtonDB"), menyoroti kompatibilitas sebagai perhatian yang dominan, diikuti oleh sebagian besar masalah lain-lain.

3.2 Analisis Sentimen Pengguna

Analisis sentimen mengungkapkan distribusi sentimen pengguna terhadap kompatibilitas Proton sebagai berikut:

Sentimen	Jumlah Laporan	Persentase
Positif	2253	48.9%
Negatif	1282	27.8%
Netral	1077	23.4%

- * **Positif (48.9%):** Hampir setengah dari ulasan mengungkapkan kepuasan, kemungkinan karena keberhasilan menjalankan game di Proton. Ini adalah indikasi kuat dari efektivitas Proton.
- * **Negatif (27.8%):** Sebagian besar laporan ketidakpuasan, seringkali terkait dengan masalah teknis atau kinerja yang buruk. Ini menunjukkan area yang perlu ditingkatkan.
- * **Netral (23.4%):** Pengguna memberikan umpan balik faktual atau ambigu tanpa nada positif atau negatif yang kuat. Ulasan ini mungkin berisi informasi berharga tentang masalah teknis yang spesifik.

Skor sentimen keseluruhan, yang dihitung sebagai skor gabungan rata-rata dari VADER, adalah **0.126**, yang menunjukkan sentimen rata-rata yang sedikit positif (skor positif > 0.05, netral -0.05 hingga 0.05, negatif < -0.05). Distribusi ini divisualisasikan dalam diagram lingkaran ("Sentimen Pengguna terhadap Kompatibilitas Proton" dalam laporan asli), yang menunjukkan mayoritas umpan balik positif yang jelas.

3.3 Frekuensi Istilah Terkait Kinerja

Istilah terkait kinerja diekstraksi dan dihitung untuk mengidentifikasi masalah umum:

Istilah	Frekuensi
Run	366
Performance	338
Fps	304
Lag	65
Smooth	65
Slow	58
Stutter	40
Fast	17

- * **'Run' (366 kali):** Menunjukkan fungsionalitas umum atau eksekusi game. Ini adalah istilah yang paling mendasar, yang mencerminkan apakah game dapat dijalankan sama sekali.
- * **'Performance' (338 kali):** Mencerminkan fokus pengguna pada kecepatan dan efisiensi game. Ini adalah istilah yang lebih umum daripada metrik kinerja yang spesifik.
- * **'Fps' (304 kali):** Metrik kunci untuk menilai kelancaran game, seringkali terkait dengan masalah *frame rate*. Ini adalah indikator kinerja yang sangat spesifik.
- * **'Lag' (65 kali):** Menunjukkan keterlambatan atau *stuttering* dalam *gameplay*.
- * **'Smooth' (65 kali):** Pengguna melaporkan *gameplay* yang lancar, seringkali secara positif.
- * **'Slow' (58 kali):** Menggambarkan kinerja yang buruk atau keterlambatan.
- * **'Stutter' (40 kali):** Mengacu pada gangguan *frame rate* atau *choppiness*.
- * **'Fast' (17 kali):** Lebih jarang, menunjukkan kinerja yang baik.

Frekuensi ini, divisualisasikan dalam diagram batang ("Frekuensi Kata Terkait Performa Game"), menggarisbawahi "run," "performance," dan "fps" sebagai aspek kinerja yang paling banyak dibahas, menunjukkan bahwa pengguna memprioritaskan eksekusi game dan *frame rate*.

3.4 Penggunaan Bahasa Berdasarkan Sentimen

Rata-rata penggunaan jenis kata (kata benda, kata kerja, kata sifat, kata keterangan) bervariasi berdasarkan sentimen:

Sentimen	Kata Benda	Kata Kerja	Kata Sifat	Kata Keterangan
Negatif	7.1	1.8	2.6	1.1
Netral	3.4	0.8	1.1	0.5
Positif	6.4	1.5	2.4	1.3

* Ulasan Negatif:

- Kata Benda: 7.1 – Penggunaan kata benda tertinggi, kemungkinan karena deskripsi masalah yang terperinci.
- Kata Kerja: 1.8 – Penggunaan kata kerja sedang, yang mencerminkan tindakan seperti "*crashes*" atau "*fails*."
- Kata Sifat: 2.6 – Penggunaan kata sifat tinggi, seringkali menggambarkan masalah (misalnya, "*unplayable*", "*slow*").
- Kata Keterangan: 1.1 – Penggunaan kata keterangan sedang, seperti "*completely*" atau "*still*."

* Ulasan Netral:

- Kata Benda: 3.4 – Penggunaan kata benda terendah, yang mencerminkan laporan faktual yang ringkas.

- Kata Kerja: 0.8 – Penggunaan kata kerja terendah, menunjukkan lebih sedikit deskripsi tindakan.
- Kata Sifat: 1.1 – Penggunaan kata sifat terendah, menunjukkan bahasa yang kurang deskriptif.
- Kata Keterangan: 0.5 – Penggunaan kata keterangan terendah, konsisten dengan nada netral.

★ **Ulasan Positif:**

- Kata Benda: 6.4 – Penggunaan kata benda tinggi, seringkali menyebutkan game atau fitur secara positif.
- Kata Kerja: 1.5 – Penggunaan kata kerja sedang, seperti "*works*" atau "*runs*."
- Kata Sifat: 2.4 – Penggunaan kata sifat tinggi, menggambarkan pengalaman positif (misalnya, "*perfect*", "*great*").
- Kata Keterangan: 1.3 – Penggunaan kata keterangan tertinggi, meningkatkan deskripsi positif (misalnya, "*flawlessly*", "*perfectly*").

Variasi ini, divisualisasikan dalam diagram batang ("Penggunaan Kata berdasarkan Sentimen Review"), menunjukkan bahwa ulasan negatif adalah yang paling deskriptif, sedangkan ulasan netral adalah yang paling ringkas.

3.5 Kata-kata yang Paling Umum Berdasarkan Sentimen

Kata-kata yang paling sering muncul di setiap kategori sentimen memberikan wawasan tentang fokus pengguna:

Sentimen	Kata-kata Paling Umum
Positif	game (1344), works (689), runs (636), perfectly (608), fine (388)
Negatif	game (845), issues (229), works (201), runs (198), error (192)
Netral	game (343), works (165), start (138), crashes (134), launch (124)

★ **Ulasan Positif:**

- "Game": 1.344 kali – Fokus utama ulasan.
- "Works": 689 kali – Menunjukkan kompatibilitas yang berhasil.
- "Runs": 636 kali – Menunjukkan eksekusi yang lancar.
- "Perfectly": 608 kali – Menyoroti kinerja yang sempurna.
- "Fine": 388 kali – Menunjukkan operasi yang memuaskan.

★ **Ulasan Negatif:**

- "Game": 845 kali – Masih menjadi fokus utama, tetapi dengan konotasi negatif.
- "Issues": 229 kali – Mencerminkan masalah yang dihadapi.

- "Works": 201 kali – Seringkali dinegasikan (misalnya, "doesn't work").
- "Runs": 198 kali – Sering dipasangkan dengan hasil negatif (misalnya, "runs poorly").
- "Error": 192 kali – Menunjukkan kegagalan teknis.

* **Ulasan Netral:**

- "Game": 343 kali – Umum di semua sentimen.
- "Works": 165 kali – Pernyataan faktual tentang fungsionalitas.
- "Start": 138 kali – Sering menggambarkan upaya peluncuran game awal.
- "Crashes": 134 kali – Pelaporan netral tentang masalah.
- "Launch": 124 kali – Mengacu pada proses inisiasi game.

Frekuensi kata-kata ini, divisualisasikan dalam *word cloud* ("Kata-kata Umum dalam Review Positive/Negative/Neutral"), menyoroti fokus positif pada fungsionalitas, fokus negatif pada masalah, dan fokus netral pada pelaporan faktual.

3.6 Statistik Panjang Ulasan Berdasarkan Kategori

Panjang rata-rata dan rentang ulasan (dalam kata) berdasarkan kategori masalah memberikan wawasan tentang keterlibatan pengguna:

Kategori Masalah	Panjang Rata-rata	Rentang
Bug	9.4	1.0 - 111.0
Bug, Kompatibilitas	20.8	1.0 - 269.0
Kompatibilitas	10.7	1.0 - 111.0
Lainnya	5.5	1.0 - 81.0
Kinerja	9.6	1.0 - 41.0
Kinerja, Bug	17.6	2.0 - 82.0
Kinerja, Bug, Kompatibilitas	32.6	3.0 - 173.0
Kinerja, Kompatibilitas	15.5	2.0 - 156.0

* **Bug:**

- Panjang rata-rata: 9.4 kata
- Rentang: 1.0 hingga 111.0 kata – Laporan singkat hingga cukup terperinci.

* **Bug, Kompatibilitas:**

- Panjang rata-rata: 20.8 kata
- Rentang: 1.0 hingga 269.0 kata – Lebih panjang karena masalah gabungan.

- * **Kompatibilitas:**
 - Panjang rata-rata: 10.7 kata
 - Rentang: 1.0 hingga 111.0 kata – Cukup ringkas.
- * **Lainnya:**
 - Panjang rata-rata: 5.5 kata
 - Rentang: 1.0 hingga 81.0 kata – Terpendek, mencerminkan masalah lain-lain atau tidak jelas.
- * **Kinerja:**
 - Panjang rata-rata: 9.6 kata
 - Rentang: 1.0 hingga 41.0 kata – Laporan yang relatif singkat.
- * **Kinerja, Bug:**
 - Panjang rata-rata: 17.6 kata
 - Rentang: 2.0 hingga 82.0 kata – Cukup terperinci.
- * **Kinerja, Bug, Kompatibilitas:**
 - Panjang rata-rata: 32.6 kata
 - Rentang: 3.0 hingga 173.0 kata – Terpanjang, mencerminkan masalah yang kompleks.
- * **Kinerja, Kompatibilitas:**
 - Panjang rata-rata: 15.5 kata
 - Rentang: 2.0 hingga 156.0 kata – Laporan yang cukup panjang.

Distribusi ini, divisualisasikan dalam *box plot* ("Panjang Review berdasarkan Kategori Masalah"), menunjukkan bahwa ulasan yang membahas berbagai masalah (misalnya, kinerja, bug, kompatibilitas) cenderung lebih panjang, menunjukkan bahwa pengguna memberikan detail lebih lanjut untuk masalah yang kompleks.

3.7 Korelasi Sentimen dan Panjang Ulasan

Korelasi antara panjang ulasan (jumlah kata) dan skor sentimen (skor gabungan) adalah **0.030**, korelasi positif yang sangat lemah. Ini menunjukkan bahwa panjang ulasan memiliki dampak minimal pada sentimen, dengan ulasan yang lebih panjang belum tentu lebih positif atau negatif. *Scatter plot* ("Hubungan Panjang Review dengan Sentimen") menunjukkan distribusi skor sentimen yang luas di semua panjang ulasan, tanpa hubungan linier yang kuat.

3.8 Masalah Teknis yang Paling Sering Dilaporkan

Masalah teknis yang paling sering dilaporkan, yang diidentifikasi melalui analisis *bigram*, meliputi:

Bigram	Frekuensi
Black screen	170
Game crashes	79
Game crash	23
Game crashed	23
Crash game	9
Black screens	5
Game crashing	5
Proton get	1
Crash gameplay	1

- * **'Black screen' (170 kali):** Masalah visual yang umum, seringkali mengindikasikan masalah kompatibilitas atau driver.
- * **'Game crashes' (79 kali):** *Crash* yang sering mengganggu *gameplay*.
- * **'Game crash' (23 kali):** Mirip dengan "Game Crashes" namun dengan variasi frasa.
- * **'Game crashed' (23 kali):** Laporan *crash* dalam bentuk lampau.
- * **'Crash game' (9 kali):** Frasa masalah *crash* yang kurang umum.
- * **'Black screens' (5 kali):** Bentuk jamak dari masalah layar hitam.
- * **'Game crashing' (5 kali):** Laporan *crash* yang sedang berlangsung.
- * **'Proton Get' (1 kali):** Sebutan yang jarang, dan kemungkinan frase yang tidak lengkap.
- * **'Crash Gameplay' (1 kali):** Sangat jarang, laporan yang cukup spesifik.

Distribusi ini, divisualisasikan dalam diagram batang ("Masalah Teknis yang Sering Dilaporkan"), menekankan "black screen" dan "game crashes" sebagai tantangan teknis utama yang dihadapi pengguna.

3.9 Statistik Umum

Statistik keseluruhan untuk dataset meliputi:

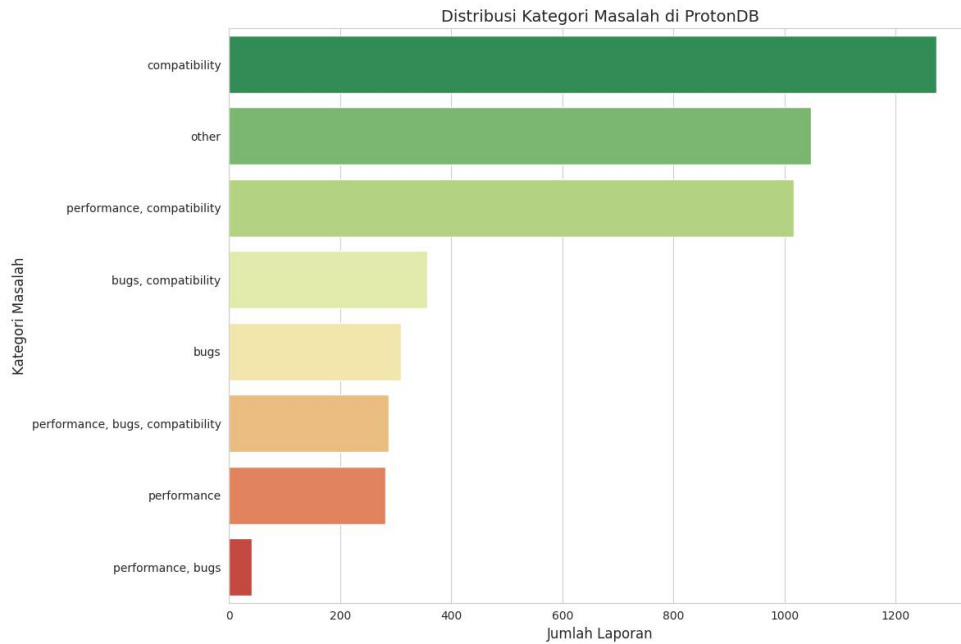
- * **Total ulasan yang dianalisis:** 4.612
- * **Rata-rata kata per ulasan:** 12.6 – Menunjukkan umpan balik pengguna yang ringkas.
- * **Rata-rata kalimat per ulasan:** 2.0 – Menunjukkan laporan yang singkat dan terfokus.
- * **Rata-rata keanekaragaman leksikal:** 0.964 – Keanekaragaman tinggi, menunjukkan penggunaan kosakata yang bervariasi.

Metrik ini memberikan pemahaman dasar tentang struktur dataset dan keterlibatan pengguna.

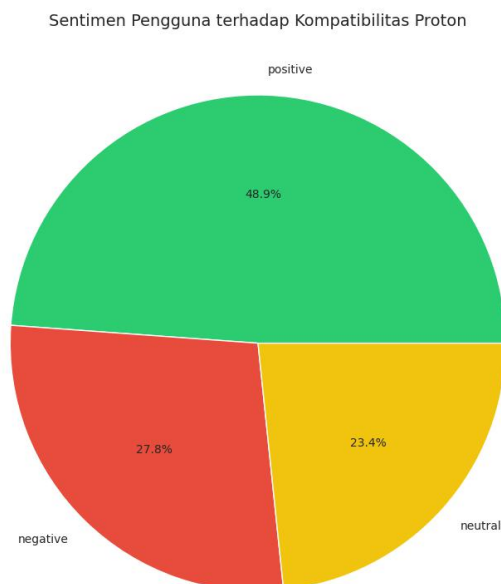
4. Visualisasi

Laporan ini menyertakan beberapa visualisasi untuk melengkapi data numerik:

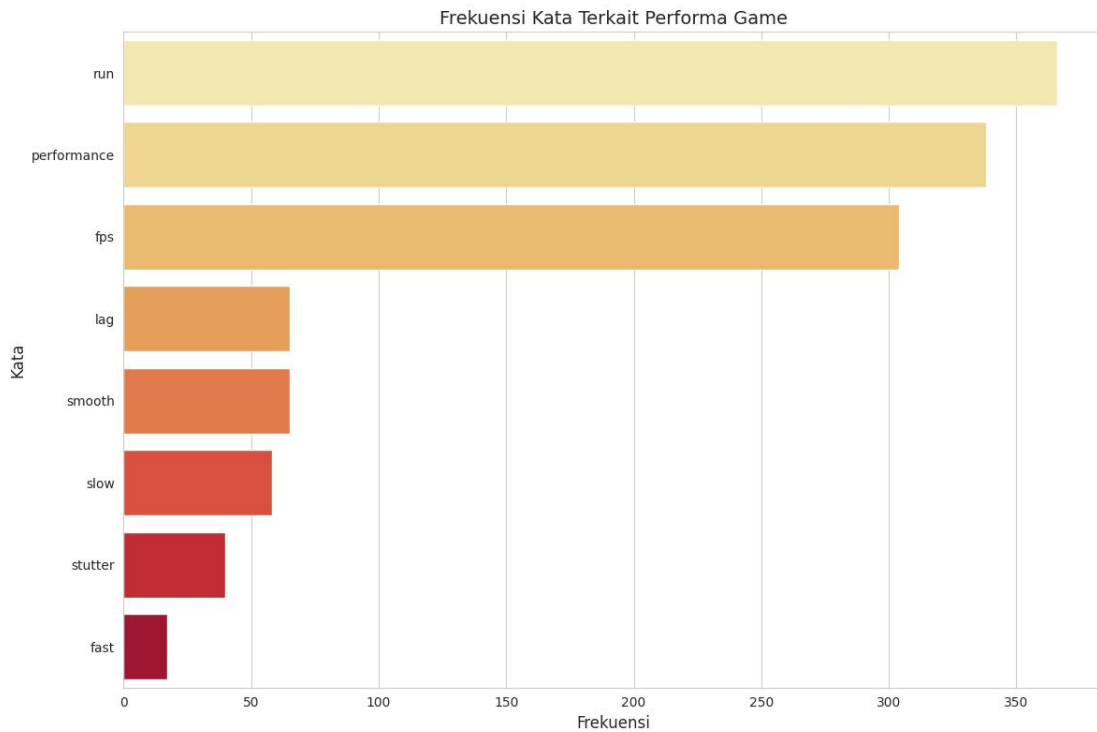
- * **Diagram Batang: Distribusi Kategori Masalah** – Menunjukkan frekuensi setiap kategori masalah, dengan *compatibility* memimpin.



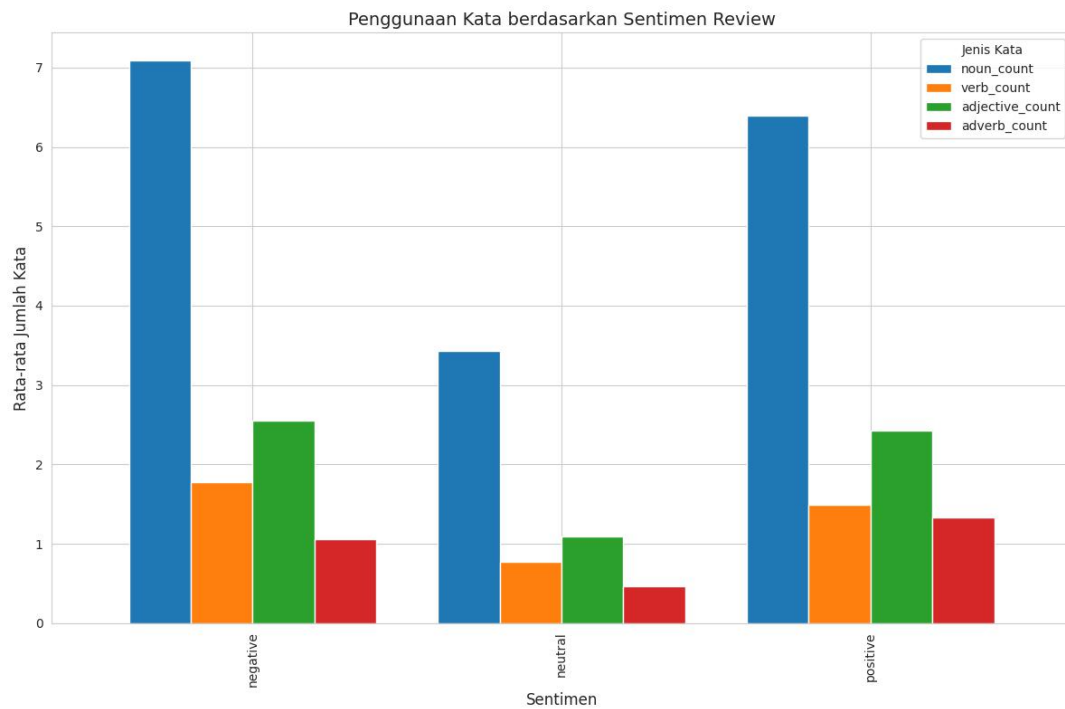
- * **Diagram Lingkaran: Distribusi Sentimen Pengguna** – Menggambarkan pembagian sentimen positif 48,9%, negatif 27,8%, dan netral 23,4%.



- * **Diagram Batang: Frekuensi Istilah Terkait Kinerja – Menyoroti "run," "performance," dan "fps" sebagai istilah teratas.**



- * **Diagram Batang: Penggunaan Bahasa Berdasarkan Sentimen – Membandingkan penggunaan kata benda, kata kerja, kata sifat, dan kata keterangan di seluruh sentimen.**



-
- Panjang Review berdasarkan Kategori Masalah**
- This box plot displays the distribution of review lengths (Jumlah Kata) across eight categories of problems. The y-axis represents the number of words, ranging from 0 to 250. The x-axis lists the categories: performance_compatibility, bugs_compatibility, compatibility, other, performance_bugs_compatibility, bugs, performance_bugs, and performance. Each category is represented by a colored box plot with individual data points overlaid. The 'performance_bugs_compatibility' category shows the highest median review length, while the 'other' category shows the lowest. Outliers are visible for several categories, particularly 'performance_compatibility' and 'bugs_compatibility'.
- | Kategori Masalah | Median (Jumlah Kata) | Q1 (Jumlah Kata) | Q3 (Jumlah Kata) | Min (Jumlah Kata) | Max (Jumlah Kata) | Outliers (Jumlah Kata) |
|--------------------------------|----------------------|------------------|------------------|-------------------|-------------------|------------------------|
| performance_compatibility | ~10 | ~5 | ~20 | ~0 | ~45 | ~120, ~155 |
| bugs_compatibility | ~15 | ~10 | ~25 | ~0 | ~50 | ~200, ~265 |
| compatibility | ~10 | ~5 | ~15 | ~0 | ~30 | ~110 |
| other | ~5 | ~2 | ~8 | ~0 | ~15 | ~80 |
| performance_bugs_compatibility | ~30 | ~15 | ~45 | ~0 | ~90 | ~145, ~170 |
| bugs | ~10 | ~5 | ~15 | ~0 | ~25 | ~55, ~60, ~110 |
| performance_bugs | ~15 | ~10 | ~20 | ~0 | ~35 | ~60, ~65, ~80 |
| performance | ~10 | ~5 | ~15 | ~0 | ~25 | ~40, ~45 |

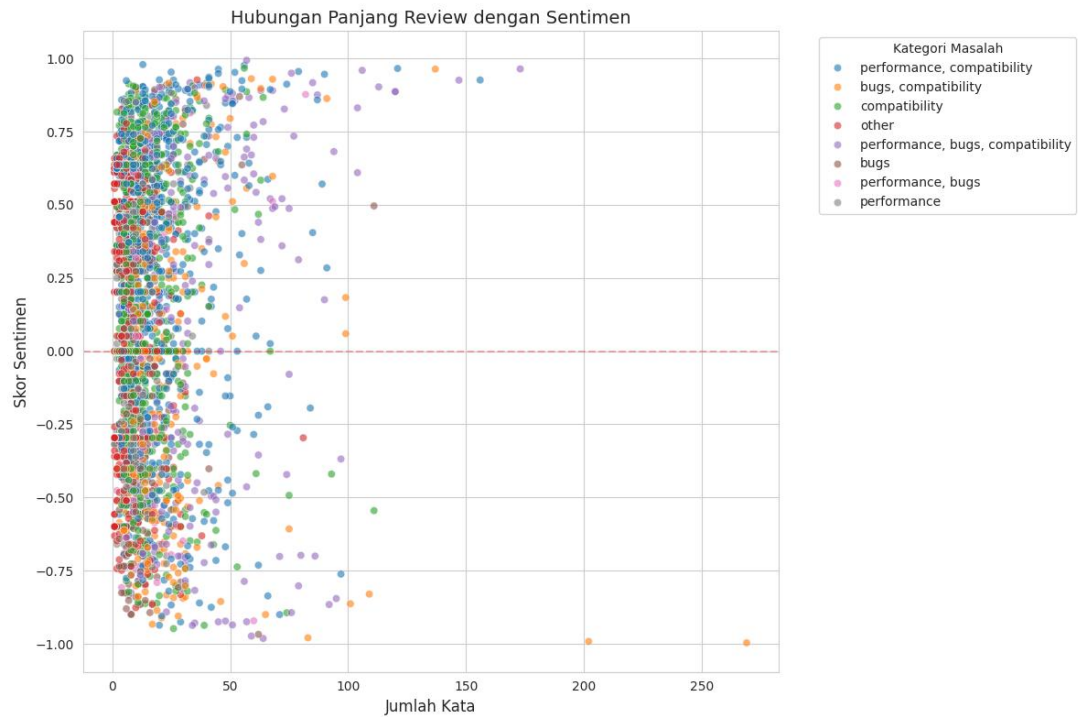
- [illegible]

word cloud positive

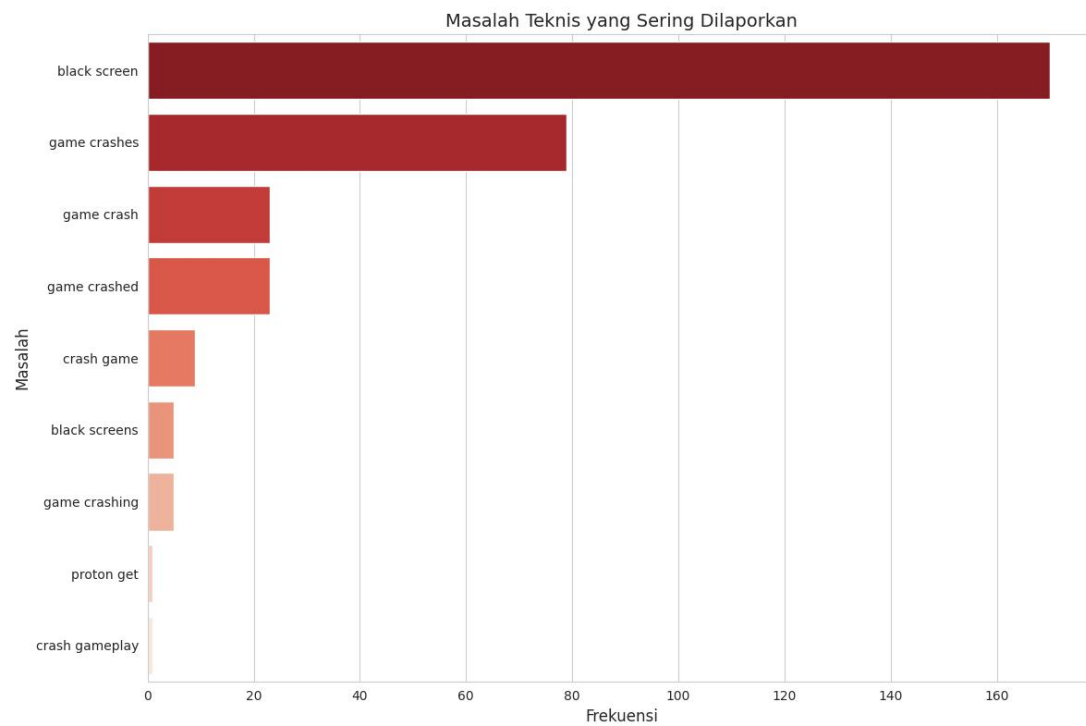
[illegible]

word cloud neutral

- * **Scatter Plot: Sentimen vs. Panjang Ulasan** – Memeriksa hubungan antara jumlah kata dan skor sentimen.



- * **Diagram Batang: Masalah Teknis yang Paling Sering Dilaporkan** – Menekankan "black screen" dan "game crashes" sebagai masalah utama.



5. Diskusi dan Implikasi

5.1 Sintesis Temuan Utama dan Maknanya

Hasil analisis ini memberikan gambaran yang kaya dan multi-faceted tentang pengalaman pengguna ProtonDB. Dominasi sentimen positif (48,9%) adalah indikasi kuat bahwa Proton, secara keseluruhan, berhasil memenuhi tujuannya sebagai jembatan antara game Windows dan ekosistem Linux. Ini merupakan pencapaian signifikan, mengingat kompleksitas teknis yang terlibat dalam emulasi dan *porting* game. Namun, sentimen negatif yang cukup besar (27,8%) dan prevalensi masalah teknis tertentu, seperti "black screen" dan "game crashes", menggarisbawahi bahwa masih ada ruang signifikan untuk perbaikan.

Frekuensi tinggi istilah-istilah yang berhubungan dengan kinerja ("run," "performance," "fps") menegaskan kembali apa yang sudah lama diketahui dalam komunitas *gaming*: kinerja adalah raja. *Gamer* sangat peduli dengan kelancaran, *frame rate* yang stabil, dan responsivitas *gameplay*. Bahkan jika sebuah game *kompatibel* dalam arti dapat dijalankan, pengalaman yang buruk karena *lag*, *stutter*, atau *frame rate* yang rendah akan sangat mengurangi kepuasan pengguna.

Perbedaan pola penggunaan bahasa antara ulasan positif, negatif, dan netral juga memberikan wawasan menarik. Ulasan negatif yang cenderung lebih deskriptif (dengan penggunaan kata benda dan kata sifat yang lebih tinggi) menunjukkan bahwa pengguna yang mengalami masalah lebih termotivasi untuk memberikan detail, mungkin dengan harapan bahwa informasi tersebut akan membantu pengembang dalam memecahkan masalah. Sebaliknya, ulasan positif cenderung lebih ringkas dan fokus pada hasil akhir ("works," "runs perfectly").

Korelasi yang sangat lemah antara panjang ulasan dan sentimen adalah temuan yang agak mengejutkan. Intuisi awal mungkin menyarankan bahwa ulasan yang lebih panjang cenderung lebih negatif (karena pengguna meluangkan waktu untuk menjelaskan masalah secara rinci) atau lebih positif (karena pengguna antusias tentang game tersebut). Namun, data menunjukkan bahwa panjang ulasan lebih merupakan fungsi dari kompleksitas masalah yang dihadapi dan gaya komunikasi individu pengguna, daripada sentimen yang mendasarinya.

5.2 Implikasi

- * **Untuk Pengembang:** Fokus pada penyelesaian masalah "black screen" dan "game crashes", mungkin dengan meningkatkan kompatibilitas driver, mengoptimalkan versi Proton, atau meningkatkan *error logging* untuk diagnostik.
- * **Untuk *Gamer*:** Memahami bahwa sebagian besar ulasan positif dapat mendorong *gaming* Linux, tetapi waspadai potensi kendala teknis seperti *crash* atau *lag* kinerja.
- * **Untuk Pengelola Platform:** Terus menyempurnakan lapisan kompatibilitas Proton, memanfaatkan umpan balik pengguna untuk menargetkan game atau konfigurasi *hardware* tertentu dengan masalah yang diketahui.

5.3 Keterbatasan

- * **Keterbatasan Ukuran Sampel:** Analisis terbatas pada 500 game, berpotensi kehilangan tren yang lebih luas dari dataset ProtonDB lengkap.

- * **Akurasi NLP:** VADER dan kategorisasi topik berbasis kata kunci mungkin salah mengklasifikasikan beberapa ulasan, terutama umpan balik yang ambigu atau bernuansa.
- * **Bias Bahasa:** Analisis berfokus pada teks bahasa Inggris, berpotensi mengecualikan ulasan non-Inggris atau nuansa budaya.
- * **Bias Temporal:** *Snapshot* data mencerminkan umpan balik pengguna hingga tanggal ekstraksi, dan masalah atau pembaruan yang lebih baru mungkin telah muncul sejak saat itu.

6. Rekomendasi

Berdasarkan temuan tersebut, rekomendasi berikut diajukan:

2. **Prioritaskan Penyelesaian Masalah Teknis:** Atasi "black screen" dan "game crashes" melalui *patch* yang ditargetkan, pembaruan driver, dan pengujian komunitas untuk game yang terpengaruh.
3. **Tingkatkan Pengoptimalan Kinerja:** Fokus pada peningkatan *frame rate*, pengurangan *lag*, dan meminimalkan *stutter*, karena ini penting untuk kepuasan pengguna.
4. **Perluas Pengujian Kompatibilitas:** Tingkatkan pengujian untuk game yang dilaporkan di bawah "kompatibilitas" dan "kinerja, kompatibilitas" untuk memastikan dukungan Linux yang lebih luas.
5. **Tingkatkan Dokumentasi:** Berikan panduan yang lebih jelas tentang pemecahan masalah umum seperti *black screen* dan *crash*, dengan memanfaatkan *bigram* dan kata-kata yang paling sering diidentifikasi.
6. **Libatkan Pengguna:** Dorong pelaporan terperinci untuk masalah yang kompleks (misalnya, kombinasi kinerja, bug, kompatibilitas) untuk mengumpulkan umpan balik yang lebih komprehensif.
7. **Pembaruan Berkala:** Pantau dan perbarui Proton secara teratur untuk mengatasi persyaratan game yang terus berkembang dan perubahan *hardware*, mempertahankan tren sentimen positif.

7. Kesimpulan

Analisis komprehensif terhadap kurang lebih 4.000+ ulasan ProtonDB ini mengungkapkan sentimen pengguna yang secara umum positif (48,9%) terhadap kompatibilitas Proton untuk *gaming* Linux, dengan "black screen" dan "game crashes" sebagai masalah teknis yang paling mendesak. Istilah terkait kinerja seperti "run," "performance," dan "fps" mendominasi diskusi pengguna, yang mencerminkan fokus pada kelancaran *gameplay*. Ulasan yang lebih panjang seringkali membahas masalah yang kompleks dan multifaset, sementara sentimen sebagian besar tidak bergantung pada panjang ulasan. Informasi ini memberikan peta jalan untuk meningkatkan fungsionalitas Proton, mengatasi masalah pengguna, dan membina ekosistem *gaming* Linux yang kuat.