



Text Mining

ANALISIS SENTIMEN DAN TOPIK ULASAN PENGGUNA APLIKASI LIVIN' BY MANDIRI MENGGUNAKAN ORANGE DATA MINING





ANGGOTA KELOMPOK

Ratu Almas Naurah Anvilén
2106752035

Yudistira Dwi Cahya
2006530942

Amira Shohifa
2206829130

Tegar Wahyu Khisbulloh
2206082032

Nadhif Maulandra Aftaura
2220628011

Ramya Nareswari Wikantiyoso
2106751606

Marietha Asnat Nauli Sitompul
2106752413





LIST OF CONTENTS



- 1. Business Understanding**
- 2. Data Understanding**
- 3. Text Mining**
- 4. Key Findings**
- 5. Conclusion**



KELOMPOK 2 - SISTEM TEMU KEMBALI INFORMASI

BUSINESS UNDERSTANDING

PENGANTAR TEMA YANG DIPILIH

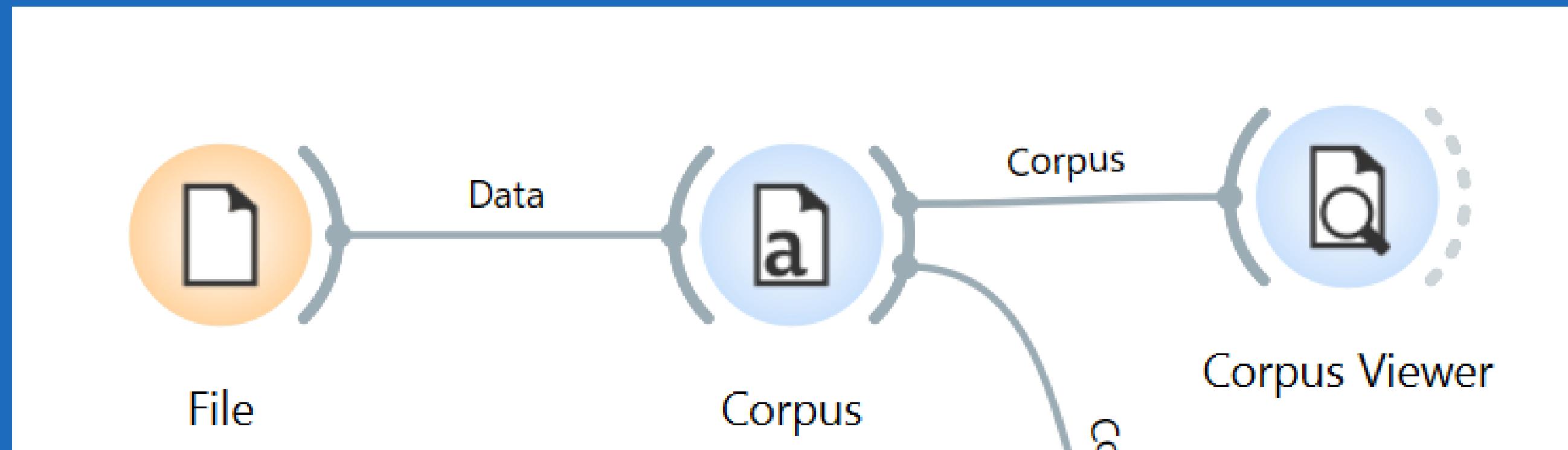
Dalam era digital yang semakin berkembang, layanan **mobile banking** menjadi kebutuhan utama bagi masyarakat dalam **mengelola keuangan secara praktis**. Salah satu aplikasi yang banyak digunakan di Indonesia adalah **Livin' by Mandiri**, yang menawarkan berbagai fitur mulai dari **transfer, pembayaran, hingga investasi**. Meskipun demikian, pengalaman pengguna terhadap aplikasi ini sangat beragam, mulai dari apresiasi hingga keluhan teknis. Oleh karena itu, penting untuk menganalisis ulasan pengguna guna memahami sentimen, emosi, serta topik yang paling sering dibicarakan. Analisis ini tidak hanya memberikan gambaran umum persepsi publik, tetapi juga **menjadi acuan bagi pengembang aplikasi untuk meningkatkan kualitas layanan**. Dengan menggunakan **Orange Data Mining**, kami melakukan **text mining** terhadap **review pengguna** untuk menggali lebih dalam makna dari setiap opini yang tertulis.



KELOMPOK 2 - SISTEM TEMU KEMBALI INFORMASI

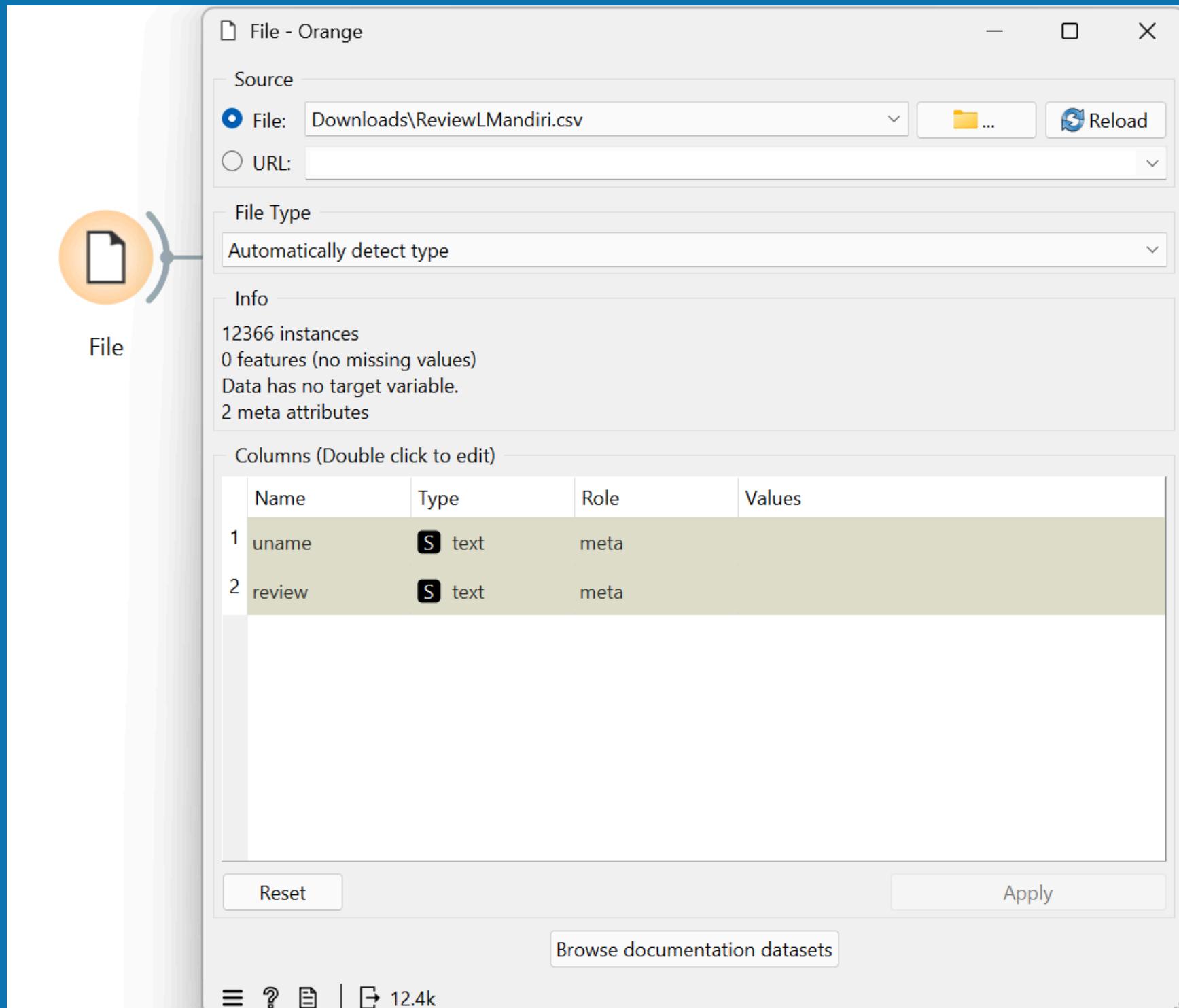
DATA UNDERSTANDING

DATA UNDERSTANDING



Tahap awal dilakukan dengan **mengimpor dataset menggunakan widget File**, kemudian diteruskan ke **Corpus** untuk **membaca teks** dalam format dokumen. Setelah itu, digunakan **Corpus Viewer** untuk **menampilkan isi ulasan** dalam bentuk mentah sehingga dapat dikenali struktur dan isi data secara menyeluruh. Langkah ini bertujuan untuk **memahami karakteristik dasar dari data teks** yang akan dianalisis

DATASET INFORMATION



Dataset yang digunakan adalah file CSV berjudul **ReviewLMandiri.csv**, yang berisi **12.366 baris** data ulasan pengguna. Data terdiri dari dua kolom:

- **uname** : nama atau identitas pengguna (sebagai meta)
- **review** : teks ulasan dari pengguna (sebagai sumber analisis utama)

Tidak terdapat nilai kosong (missing values) dan belum ada target variable, karena tujuan utama adalah **menganalisis teks berdasarkan isi ulasan, bukan klasifikasi awal**

The screenshot shows the 'Corpus Viewer - Orange' application interface. On the left, a vertical list of review IDs from 1 to 16 is displayed, with 'Mike Xun' (ID 2) selected and highlighted with a blue background. The main panel on the right shows the details for the selected review:

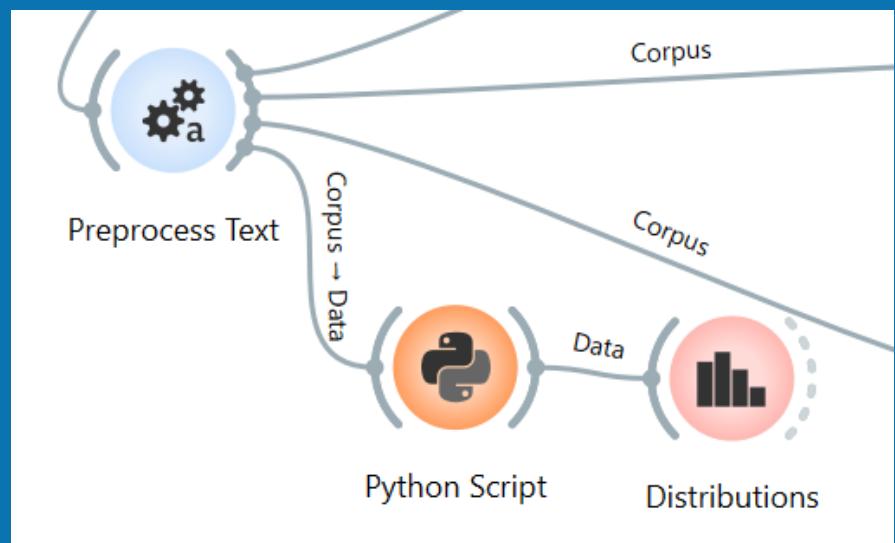
uname: Mike Xun
review: Kasih bintang 1 karena bank mandirinya yang asal kirim dan aktivasi kartu kredit tanpa pemberitahuan, tiba-tiba saja ada email notifikasi biaya tahunan kartu kredit!!! Padahal dari awal sudah tolak dan tidak pernah mengajukan kartu kredit!!!

At the bottom of the application window, there are several icons and text elements: a list icon, a question mark icon, a document icon, a right arrow icon, and the text '12.4k | 12.4k | 12.4k'.

Menggunakan **Corpus Viewer**, kita dapat melihat langsung isi dari masing-masing review yang diberikan pengguna. Salah satu contoh ulasan berbunyi :

"Kasih bintang 1 karena bank mandirinya yang asal kirim dan aktivasi kartu kredit tanpa pemberitahuan..."

Ini menunjukkan bahwa **review sangat beragam**, mulai dari **pengalaman negatif hingga positif**, dan memiliki **struktur kalimat naratif** yang bisa dianalisis lebih dalam melalui teknik text mining



The screenshot shows the 'Python Script - Orange' interface. The 'Editor' tab contains the following Python code:

```
def python_script(in_data):
    1 import pandas as pd
    2 from Orange.data import Table, Domain, ContinuousVariable
    3
    4 reviews = [str(inst['review']) for inst in in_data]
    5
    6 # Hitung panjang kalimat (jumlah kata)
    7 lengths = [len(r.split()) for r in reviews]
    8
    9 # Tambahkan ke data
    10 new_var = ContinuousVariable("review_length")
    11 domain = Domain(in_data.domain.attributes + (new_var,), in_data.domain.class_vars)
    12 in_data = Table(domain, [list(inst) + [lengths[i]] for i, inst in enumerate(in_data)])
    13 out_data = new_data
    14

    return out_data, out_learner, out_classifier, out_object
```

The 'Console' tab shows the output of running the script, including a warning message:

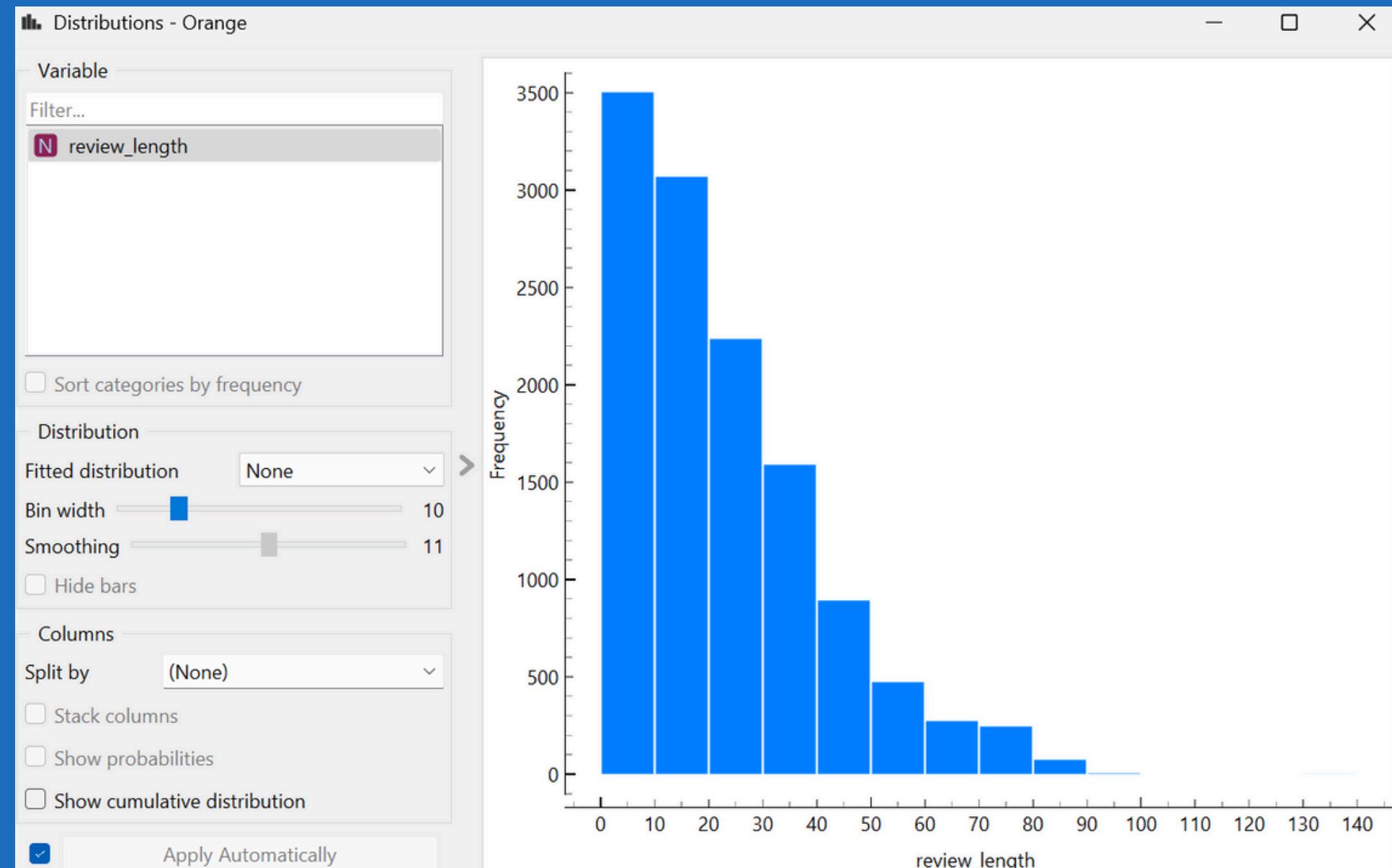
```
>>>
Running script:
<string>:13: OrangeDeprecationWarning: Direct calls to
Table's constructor are deprecated and will be removed.
Replace this call with Table.from_list
>>>
Running script:
<string>:13: OrangeDeprecationWarning: Direct calls to
Table's constructor are deprecated and will be removed.
Replace this call with Table.from_list
>>>
```

Untuk memperkaya pemahaman tentang dataset, kami menambahkan **proses penghitungan panjang review dalam jumlah kata** menggunakan widget **Python Script** di Orange. Dengan menulis skrip Python khusus, kami menghitung berapa banyak kata yang terkandung dalam setiap ulasan pengguna.

Hasil penghitungan ini kemudian **divisualisasikan** menggunakan widget **Distributions** yang menampilkan **distribusi frekuensi panjang review secara grafis**. Visualisasi ini membantu **mengidentifikasi pola umum**, yaitu sebagian besar review terdiri dari jumlah kata yang relatif pendek, sementara beberapa review memiliki panjang yang jauh lebih panjang.

Pendekatan ini penting untuk **memahami karakteristik data teks** sebelum dilakukan proses analisis lanjutan seperti ekstraksi fitur dan pemodelan

DISTRIBUSI JUMLAH KATA



Untuk memahami karakteristik ulasan secara lebih mendalam, dilakukan **analisis distribusi panjang review berdasarkan jumlah kata**. Visualisasi ini menunjukkan bahwa mayoritas ulasan pengguna memiliki **panjang kurang dari 20 kata**, dengan puncak frekuensi berada di kisaran **5–15 kata**.

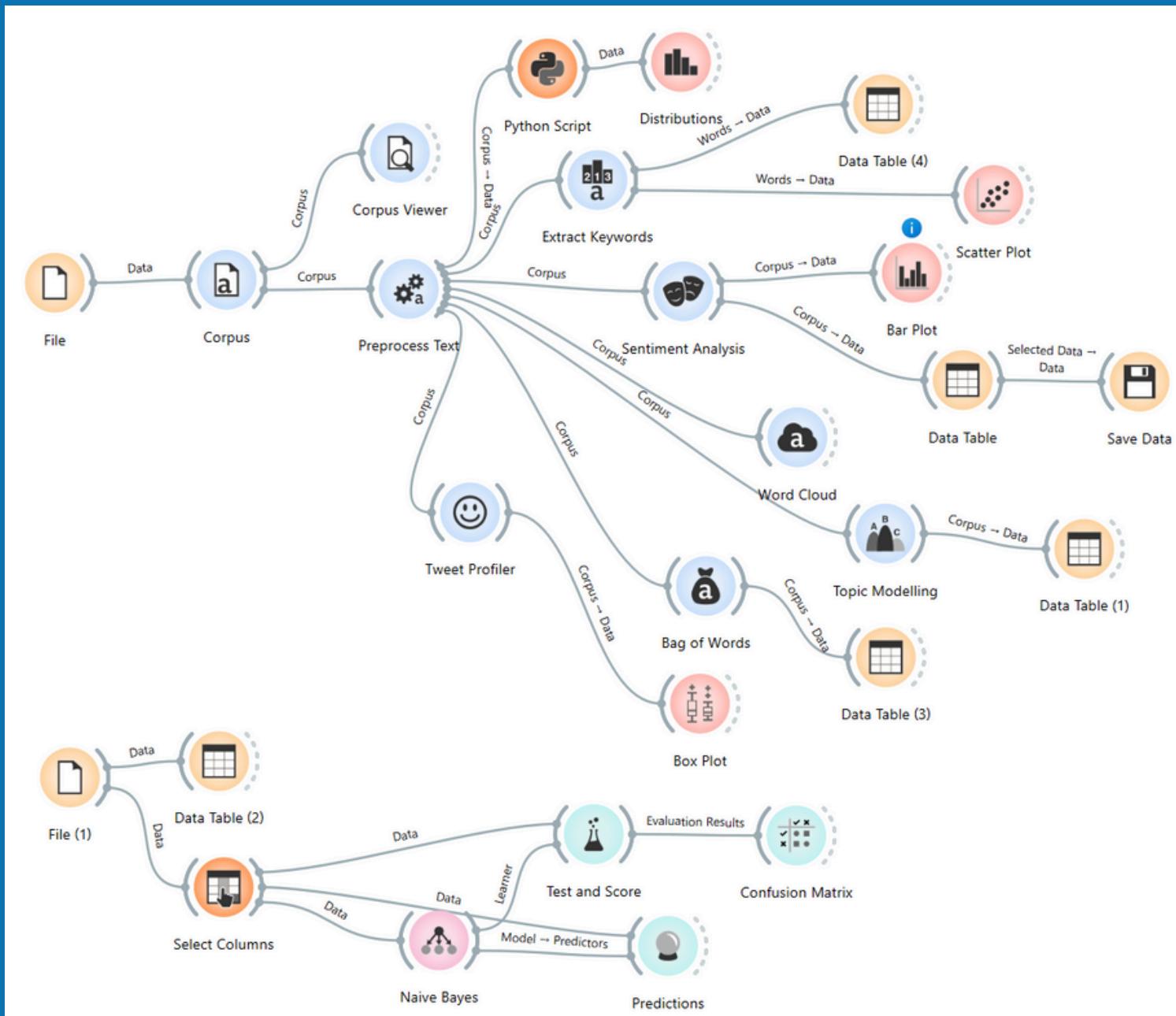
Distribusi ini memiliki pola **right-skewed**, artinya **sebagian besar review ditulis secara singkat**, namun **terdapat sejumlah kecil ulasan yang sangat panjang (lebih dari 60 kata)**. Temuan ini penting karena menunjukkan bahwa meskipun sebagian pengguna memberikan komentar singkat, ada juga yang memberikan **masukan secara rinci**, yang bisa menjadi **sumber insight** yang mendalam dalam analisis topik maupun sentimen.



KELOMPOK 2 - SISTEM TEMU KEMBALI INFORMASI

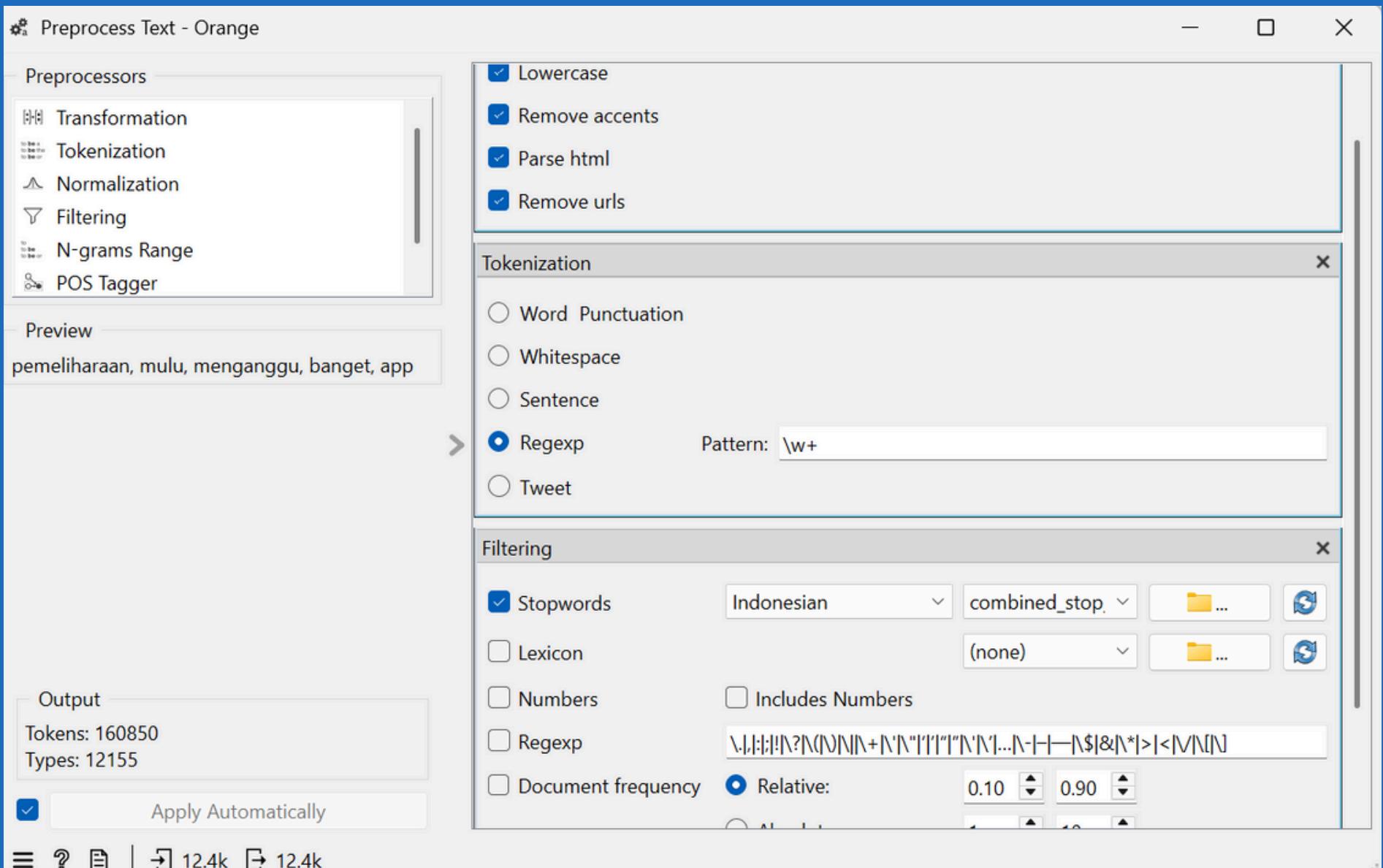
TEXT MINING

WORKFLOW TEXT-MINING



1. Dimulai dari **pengimporan data review pengguna** aplikasi Livin' by Mandiri ke dalam Orange melalui widget **File** dan **Corpus**
2. Selanjutnya, data teks **diproses menggunakan Preprocess Text** untuk **membersihkan dan menyiapkan data**, termasuk **tokenisasi, penghapusan stopwords, dan normalisasi teks**
3. Dari preprocessing, data dialirkan ke berbagai widget analisis, antara lain :
 - Extract Keywords untuk mengekstrak kata kunci utama dari ulasan
 - Sentiment Analysis untuk mengkategorikan ulasan menjadi sentimen positif, negatif, atau netral
 - Word Cloud untuk visualisasi kata yang sering muncul
 - Topic Modelling untuk mengidentifikasi topik utama dari kumpulan review
 - Tweet Profiler untuk menggali karakteristik lebih lanjut (opsional)
 - Bag of Words untuk ekstraksi fitur teks yang digunakan pada pemodelan
 - Visualisasi menggunakan Bar Plot, Scatter Plot, dan Box Plot untuk memperlihatkan distribusi dan pola data
4. Pada bagian **modeling**, dilakukan pemilihan kolom dengan **Select Columns**, diikuti penerapan algoritma **Naive Bayes** untuk klasifikasi sentimen
5. Hasil klasifikasi dievaluasi menggunakan widget **Test and Score** dan divisualisasikan dengan **Confusion Matrix** serta prediksi yang dihasilkan oleh model
6. Workflow ini memberikan gambaran komprehensif dari **proses text mining** mulai dari **persiapan data, analisis, hingga evaluasi model**

PREPROCESS TEXT

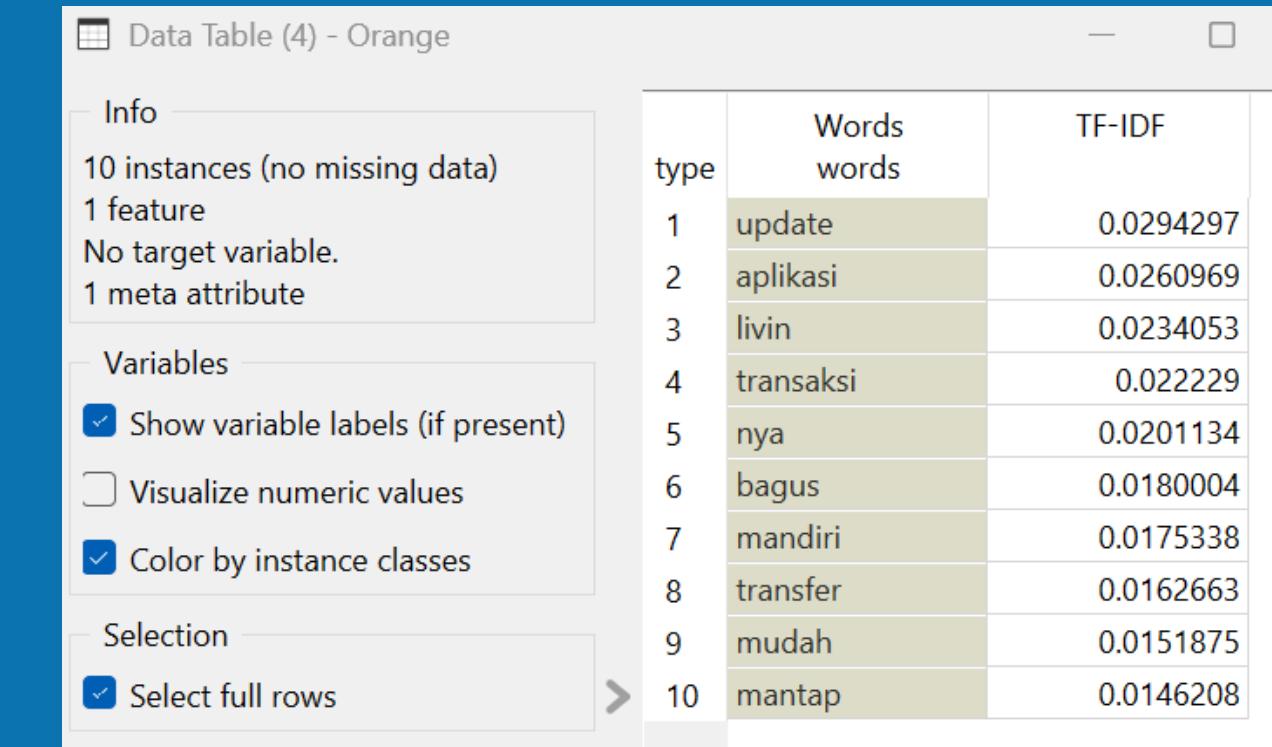
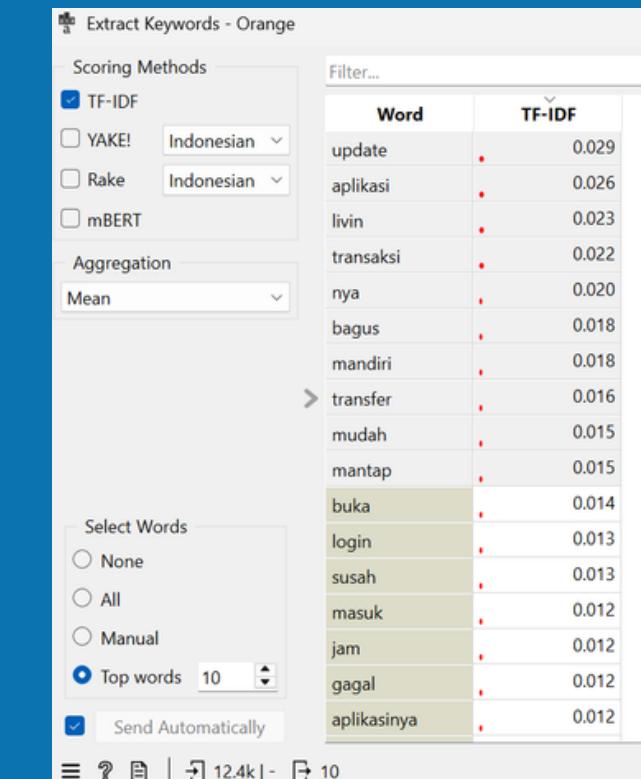
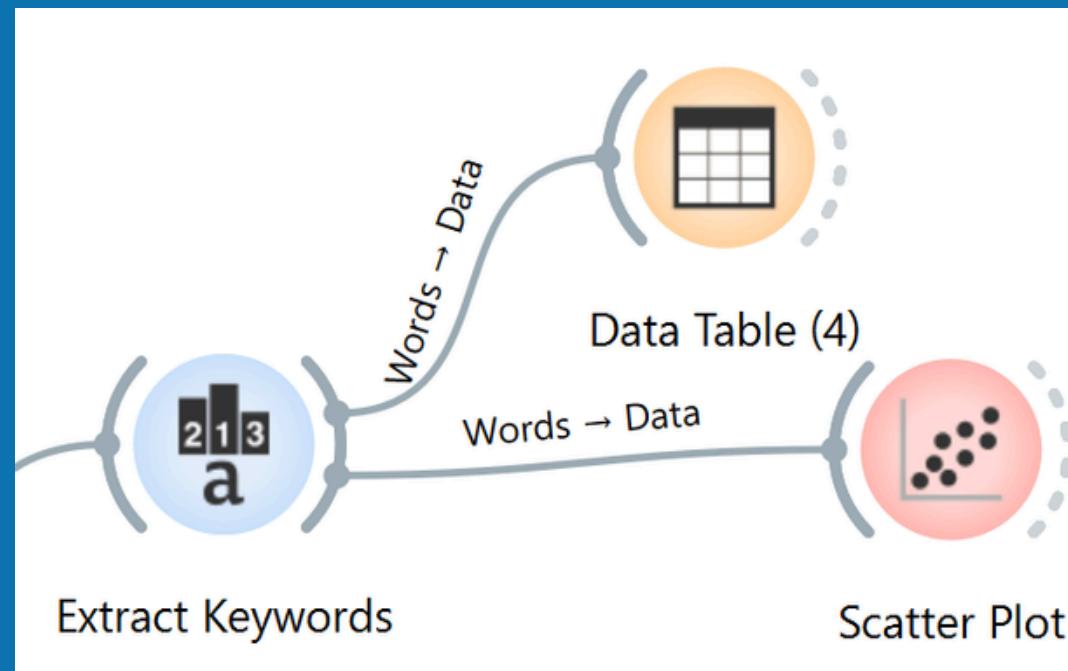


Pada tahap ini, dilakukan **pembersihan data teks** dengan beberapa proses, antara lain:

- **Mengubah** semua huruf menjadi **lowercase** agar konsisten
- **Menghapus aksen** dan **karakter HTML** yang tidak relevan
- **Menghapus URL** yang biasanya tidak berkontribusi pada makna teks
- **Tokenisasi** menggunakan pola **regexp** untuk **memisahkan kata berdasarkan huruf dan angka**
- **Penghilangan stopwords bahasa Indonesia** untuk menghilangkan kata-kata umum yang tidak penting

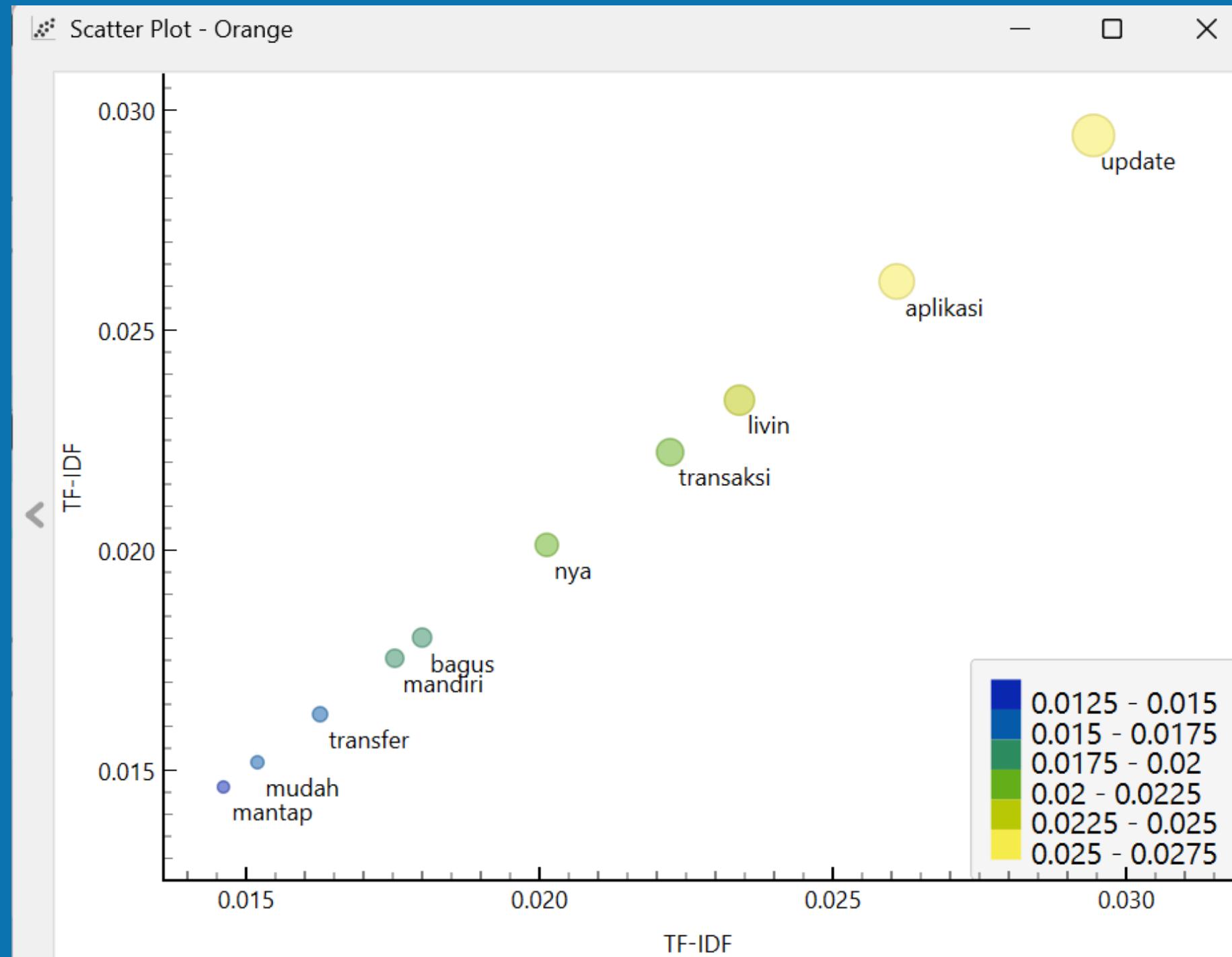
Proses ini bertujuan mempersiapkan data agar siap untuk dianalisis lebih lanjut dengan metode **text mining**

EXTRACT KEYWORDS



Proses Extract Keywords dilakukan dengan **metode TF-IDF** untuk **mengidentifikasi kata-kata kunci paling signifikan** dalam ulasan pengguna aplikasi Livin' by Mandiri. TF-IDF mengukur **seberapa penting sebuah kata** dalam dokumen relatif terhadap seluruh koleksi dokumen, sehingga **membantu menyoroti kata-kata yang paling relevan**. Kata-kata teratas seperti “**update**”, “**aplikasi**”, “**livin**”, dan “**transaksi**” mencerminkan **aspek penting** yang sering dibahas dalam ulasan. Hasil ekstraksi kata kunci ini kemudian disimpan dalam **Data Table** yang memuat **kata-kata dan skor TF-IDF-nya** secara rinci, sehingga dapat dianalisis lebih lanjut

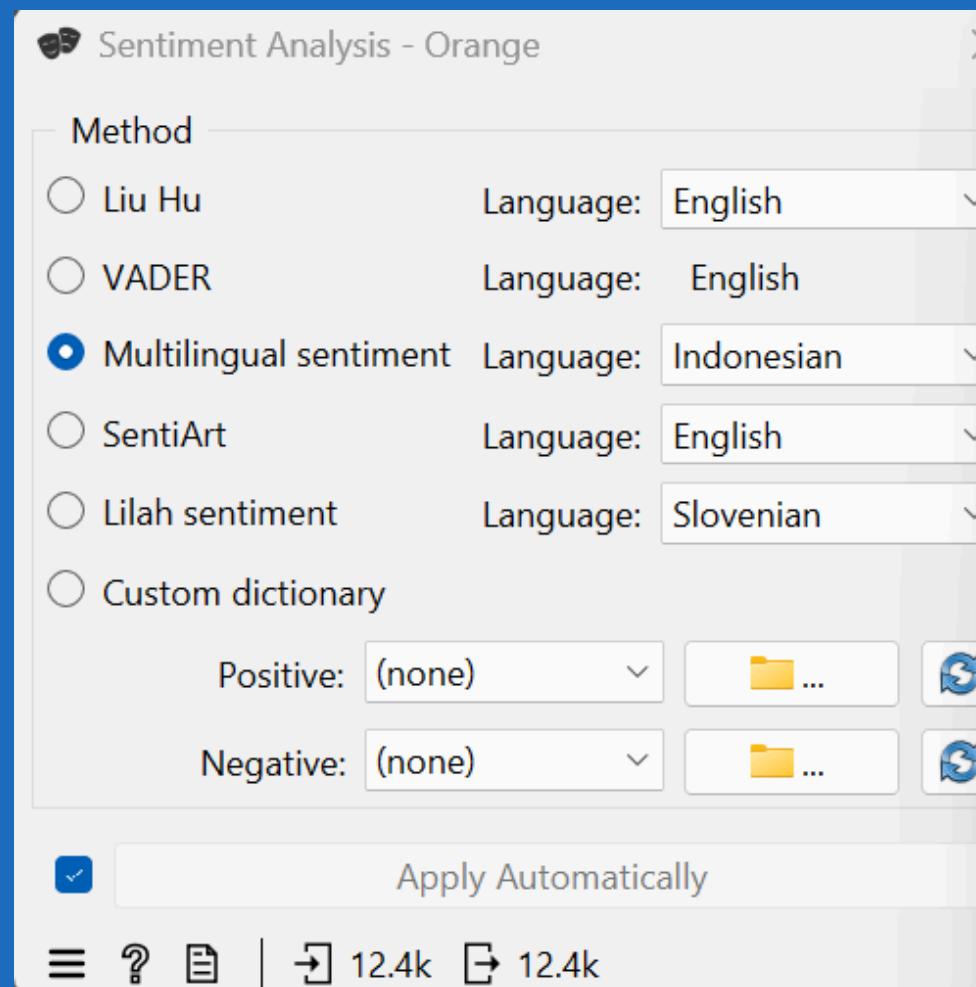
EXTRACT KEYWORDS



Untuk visualisasi, digunakan **Scatter Plot** yang memetakan kata-kata berdasarkan nilai TF-IDF-nya, dimana ukuran dan warna titik menggambarkan **bobot penting** kata tersebut

Visualisasi ini memudahkan **identifikasi kata kunci yang paling menonjol** secara intuitif dan membantu memfokuskan analisis pada **aspek-aspek utama yang dibahas pengguna**

SENTIMENT ANALYSIS

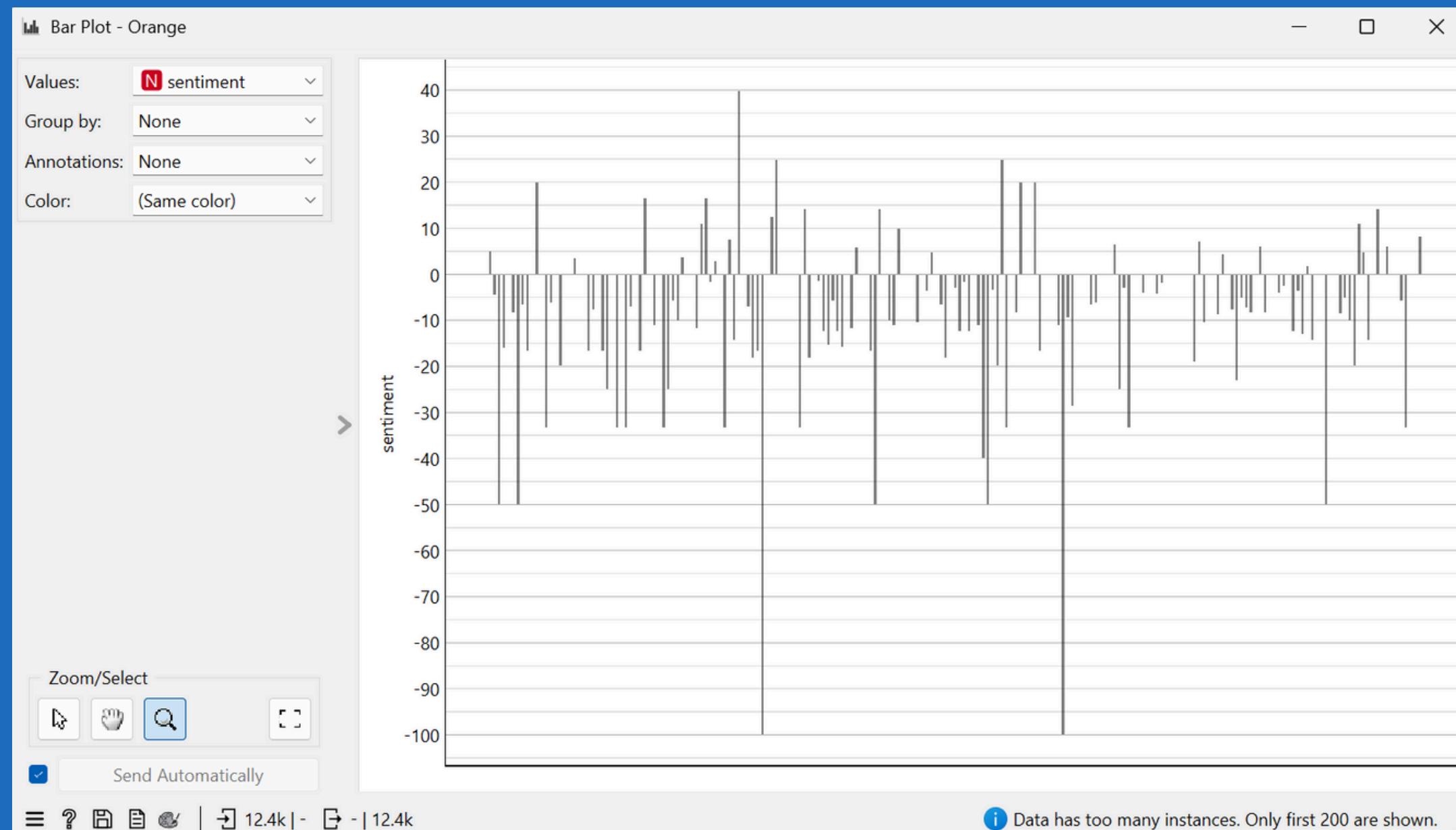


include title	uname	review	sentiment
	True	True	
60	yusuf rinaldy	Mengecewakan	-100
124	Ganjar Ikhsan H...	Sering gangguan	-100
288	Ersha Radiva	Garbage	-100
307	Agie Anugrah	busuk	-100
490	Arsenio Delano	Payah	-100
530	Atep Supriatna	Buruk	-100
544	agus wardoyo	Banyak bug	-100
583	Amane Yui	Error terus, tolong diperbaiki.	-100
991	Rizqi S Fikri	Masih banyak bug	-100
1198	Ferdiansyah Per...	Bug banyak	-100
1203	Tarsono T	sering crash sendiri.	-100
1305	Char	Ini kenapa ya kok jadi sering crash	-100
1430	Ferdiansyah Iq...	Sering error payah	-100
1523	Gilang Dayu	Kenapa sering sekali error?	-100
1883	wznic	Kenapa yg sekarang sering error ya	-100
1924	D Mahmudin	Lambat	-100
2027	dodi s	cukup lengkap	-100

include title	uname	review	sentiment
	True	True	
1356	Ray Montano	Lumayan. Tapi tidak terlalu memu...	100
1504	Wisnoe Marma...	Kurang mudah dalam penggunaan	100
1583	friska tiama putri	Terlalu banyak pembaruan	100
2058	HERMAN V	Bagus sekali.	100
2419	Rahmat Risky S	Semangat	100
2421	Budiman R Sara...	Cepat dan aman	100
2442	Henny Psrnt (h...	Sangat membantu dan mudah dig...	100
2465	Emil Lee	bagus	100
2478	Andri Dwiyana ...	Berguna	100
2526	dwi adiningrum	Memudahkan	100
2529	indra darmawan	Memudahkan	100
2544	Arif Aulya Rah...	Sangat baik dan stabil	100
2562	I Putu Gede Wir...	Sangat bagus sekali	100
2594	Arifin Sukardi	Memuaskan	100
2626	Dicky Fadlian	cukup bagus lah	100
2627	HARY SUBAYU	Super ...	100
2630	Fransisko Cantu...	Terbaik	100

Dengan menggunakan modul **Sentiment Analysis** di Orange, review diidentifikasi sebagai **sentimen positif** dan **negatif**. Contoh data menunjukkan review **negatif** berisi keluhan seperti “sering gangguan” dan “bug banyak”, sedangkan review **positif** berisi puji seperti “bagus sekali” dan “memuaskan”. Penentuan sentimen ini menggunakan metode **Multilingual Sentiment** dengan pengaturan bahasa Indonesia, yang relevan dengan isi dataset

BAR PLOT SENTIMENT ANALYSIS

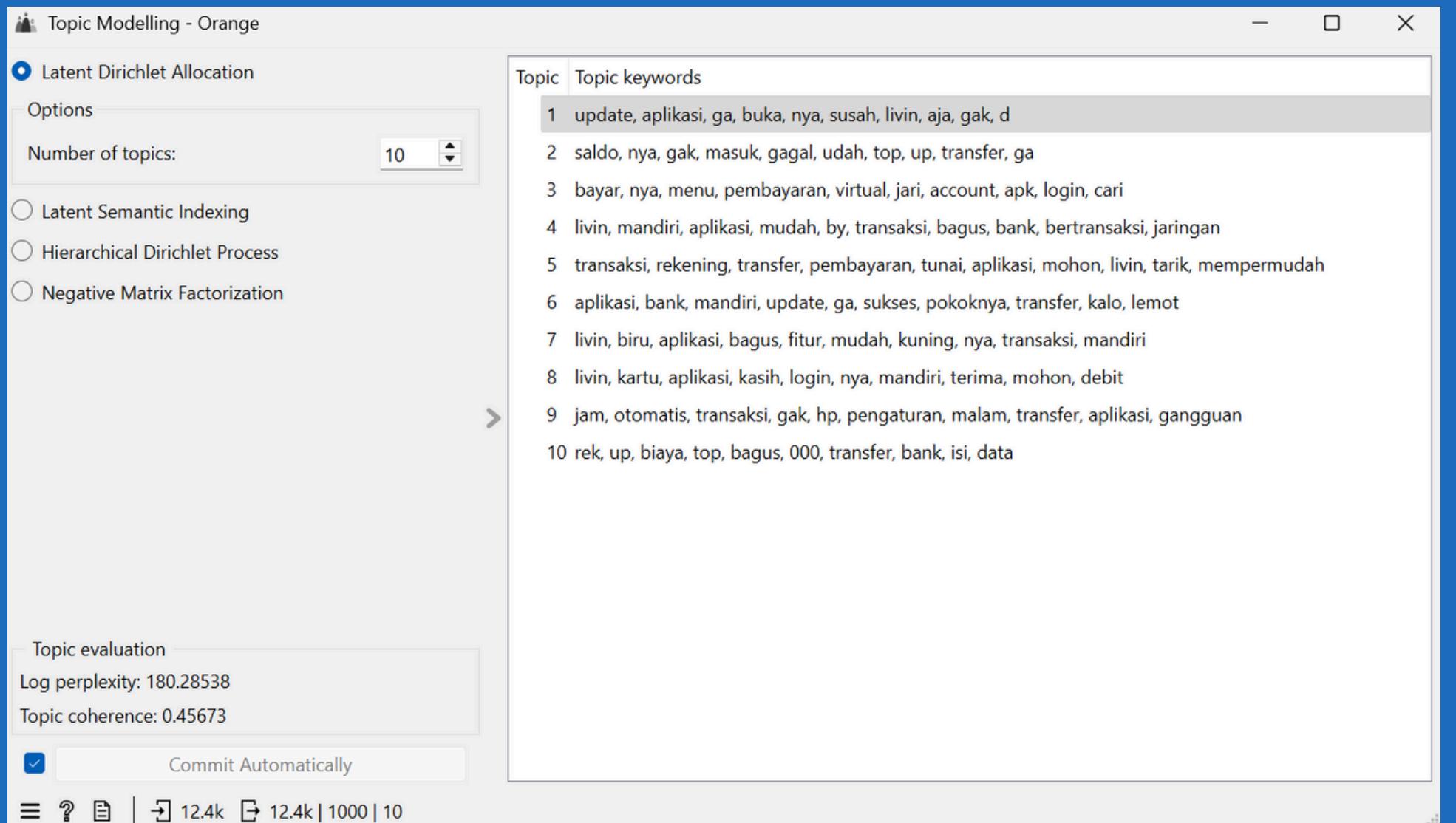


- Visualisasi **bar plot** menunjukkan **sebaran sentimen** dari seluruh review yang dianalisis
- **Sumbu Y** menunjukkan nilai sentimen dari **negatif ke positif**, sementara **sumbu X** mewakili **masing-masing review**
- Grafik ini memperlihatkan **variasi yang cukup besar** dalam sentimen, dengan **sejumlah ulasan yang sangat negatif** dan **sejumlah lainnya yang sangat positif**
- Data ini memberikan gambaran umum bagaimana pengguna merasakan aplikasi Livin' by Mandiri



- Word Cloud menampilkan **kata-kata yang paling sering muncul** dalam ulasan pengguna aplikasi Livin' by Mandiri
 - Kata-kata yang **lebih besar menunjukkan frekuensi kemunculan yang lebih tinggi**
 - Kata-kata **populer** seperti **"transaksi"**, **"aplikasi"**, **"update"**, **"livin"**, dan **"mandiri"** menunjukkan fokus utama ulasan pada aspek **fungsional aplikasi** dan **pengalaman transaksi**
 - Word Cloud ini membantu **mengidentifikasi istilah kunci** yang mendominasi percakapan pengguna secara visual dan cepat

TOPIC MODELLING RESULT



- Topic Modelling menggunakan metode **Latent Dirichlet Allocation (LDA)** untuk **mengelompokkan ulasan** menjadi beberapa **topik utama** berdasarkan **kata-kata kunci** yang sering muncul bersama
- Hasil menunjukkan **10 topik berbeda** dengan kata kunci yang mewakili tema dalam ulasan, seperti **masalah saldo, proses transfer, fitur aplikasi, dan pengalaman pengguna**
- Setiap topik mengindikasikan **area perhatian pengguna** yang berbeda-beda, yang berguna untuk analisis lebih fokus terhadap aspek-aspek layanan aplikasi

TOPIC MODELLING RESULT

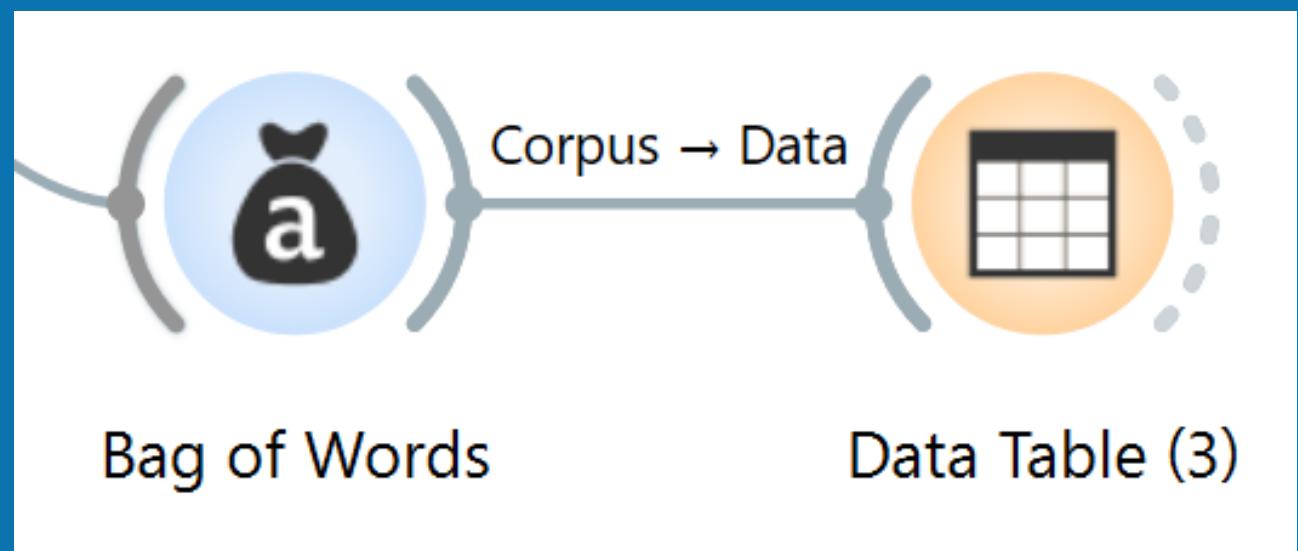
include title	uname	review True	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
			True	0.4256	0.0142886	0.264709	0.0142864
1	tabuty	Pemeliharaan ...		0.4256	0.0142886	0.264709	0.0142863
2	Mike Xun	Kasih bintang 1...		0	0	0.0632592	0
3	unknown 70	apasih anjir veri...		0	0.44839	0.516825	0
4	Mohamad Zaid	Gagal terus tra...	0.0333359	0.0333455	0.0333334	0.0333377	0.0333408
5	Lutfannisa S	Jangan lakukan...	0	0	0.448289	0	0.0600429
6	Andreas Susanto	Kenapa kalau b...	0.0923426	0.197202	0.0981805	0	0
7	Erwin Sitepu	aplikasi bank m...	0	0	0	0	0
8	Jean Nasadit	Aplikasih sampah	0.0333333	0.0333333	0.0333333	0.0333333	0.366664
9	BPK	Aplikasi freeze ...	0.174057	0	0	0.505434	0
10	Samuel Manik	Gimana sih biki...	0	0.501656	0.292669	0	0
11	Irvan Satria	Transfer Valas n...	0	0	0.608617	0	0.31864
12	Ahfad Fauzan	Agak ribet kala...	0.0166778	0.0166698	0.0166693	0.0166707	0.0166674
13	Abxyz	Ga jelas, suruh ...	0.0166738	0.315094	0.0166689	0.0166667	0.0166672

Tabel ini menampilkan **distribusi probabilitas** setiap ulasan terhadap **topik-topik** yang sudah diidentifikasi oleh model

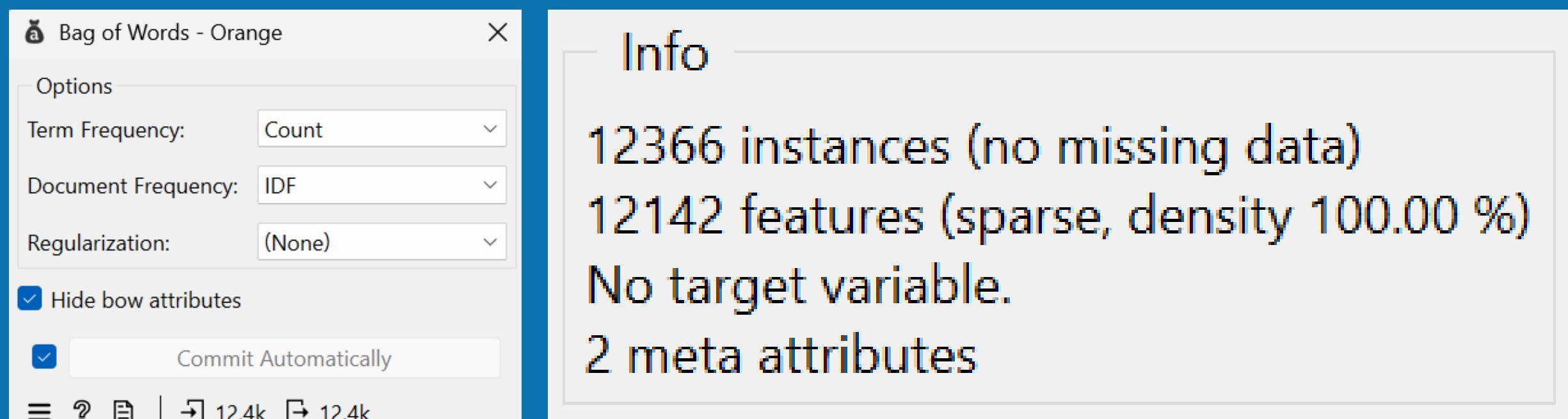
Misalnya, satu review bisa **memiliki probabilitas tinggi pada topik 1 dan rendah pada topik lainnya**

Data ini memungkinkan **pemetaan ulasan ke dalam cluster topik secara kuantitatif** untuk analisis lanjutan dan penarikan kesimpulan yang lebih mendalam

BAG OF WORDS



Pada tahap ini, digunakan metode **Bag of Words** dengan parameter **Term Frequency (Count)** dan **Document Frequency (IDF)** untuk mengekstrak fitur kata dari kumpulan review. Metode Bag of Words **mengubah teks menjadi representasi numerik** berupa **frekuensi kemunculan kata** dalam dokumen, sehingga memungkinkan **analisis kuantitatif** terhadap isi teks.



Hasil ekstraksi ditampilkan dalam bentuk **Data Table** yang memuat ribuan fitur kata (**12.142 fitur**) dari total **12.366 ulasan**, menunjukkan **tingkat kepadatan data yang tinggi** (density 100%).

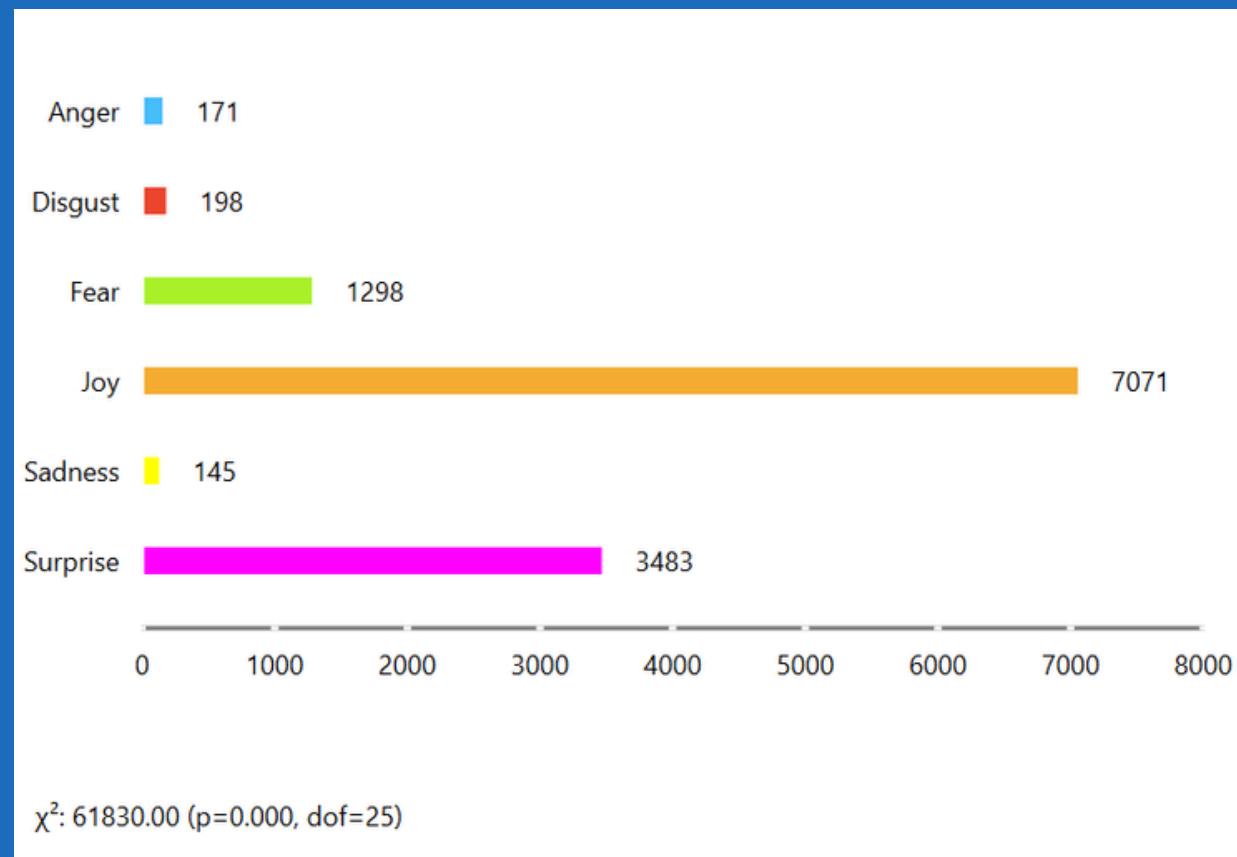
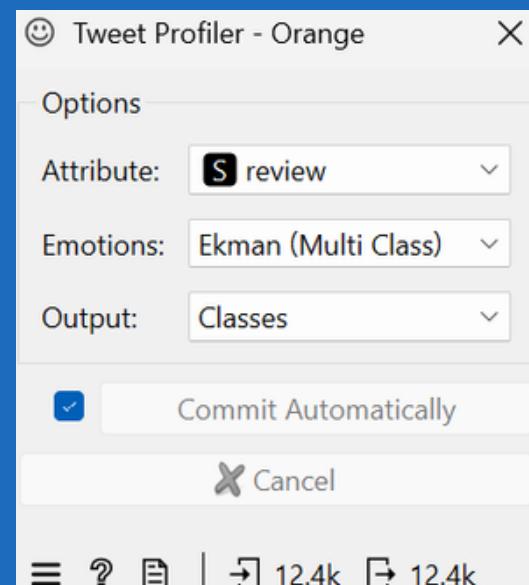
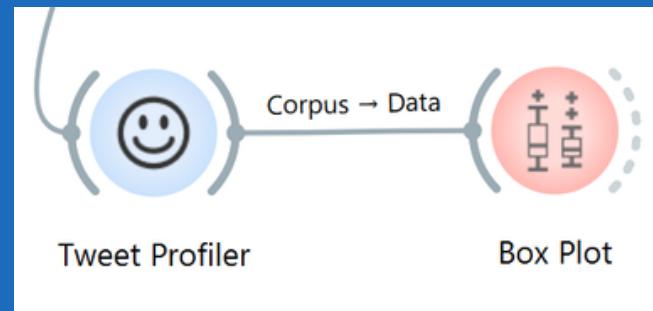
BAG OF WORDS RESULTS

bow-feature hidden include skip-normalization title	uname (1)	review (1)	...
	True	True	...
1	tabuty	Pemeliharaan mulu, menganggu banget app nya	app=3.52555, banget=2.95812, mengang...
2	Mike Xun	Kasih bintang 1 karena bank mandirinya yang asal kirim dan a...	1=3.42129, aktivasi=5.70913, bank=2.633...
3	unknown ...	apasih anjir verifikasi wajah gagal trus padahal di luar ruanga...	anjir=7.63095, apasih=7.81327, banget=2...
4	Mohama...	Gagal terus transaksi	gagal=2.77762, transaksi=1.85072
5	Lutfannis...	Jangan lakukan penggantian kartu debit mandiri secara onlin...	15=6.24465, 2024=5.55151, bank=2.6337...
6	Andreas S...	Kenapa kalau buka aplikasi, sudah disetting masukin semua d...	aplikasi=4.40922, bank=2.63373, bikin=3....
7	Erwin Site...	aplikasi bank macam apa yang offline jam 2 pagi sampai jam ...	2=3.36358, 4=4.40543, aplikasi=1.46974, ...
8	Jean Nasa...	Aplikasih sampah	aplikasih=7.63095, sampah=5.75914
9	BPK	Aplikasi freeze saat di halaman verifikasi wajah sehingga terp...	aplikasi=1.46974, app=3.52555, berpeng...
10	Samuel M...	Gimana sih bikin promo tap to pay tp gagal terus transaksiny...	bikin=6.91312, gagal=2.77762, malu=6.4...
11	Irvan Satria	Transfer Valas nya sudah tidak aktif beberapa hari? Disana ka...	aktif=5.09197, dana=3.75975, disana=8.7...
12	Ahfad Fa...	Agak ribet kalau buka tabungan via Livin	buka=2.66461, livin=1.56639, ribet=3.282...
13	Abxyz	Ga jelas, suruh buat password yang belum pernah dipake. Pa...	dipake=9.51853, password=3.83921, suru...
14	Maranath...	Password saya sudah benar malah dibilang salah	dibilang=6.16461, password=3.83921, sal...
15	Masyhud ...	Sdh 1 tahun sy menggunakan aplikasi ini & sampai sekarang ...	1=3.42129, aplikasi=1.46974, aplikasinya...
16	Ezekiel	harus banget sesi berakhir tanpa ada sepengetahuan beberapa...	banget=5.91624, bayar=3.41635, gajelas...
17	Angga Kri...	sulit dibuka. ketika di klik, seketika itu out sendiri aplikasinya	aplikasinya=2.75575, dibuka=3.51191, kli...
18	mich wirja	Aplikasi tolol ndak punya otak	aplikasi=1.46974, ndak=5.83919, otak=7...
19	Dinovryan	Apps nya sering ke detect self crashing, jadi efek ke HP self re...	apps=4.60242, bugs=6.65012, crashing=...
20	Dhinda C...	fitur kurang bagus, krn ketika aplikasi sign out, gk bisa sign in...	aplikasi=1.46974, bagus=2.4417, cabang...
21	Annisa Bila	Baru sebulan ga diisi udah diblokir	diblokir=7.02481, diisi=6.53233, sebulan...
22	Arie Unda...	Coba ganti debit online yg memamsuki masa kadaluarsa. Na...	bang=6.71466, coba=3.26573, debit=3.7...

Setiap **baris** dalam tabel merepresentasikan **satu review** dengan **nilai bobot untuk setiap kata**, yang dapat digunakan sebagai input untuk proses pemodelan dan klasifikasi lebih lanjut

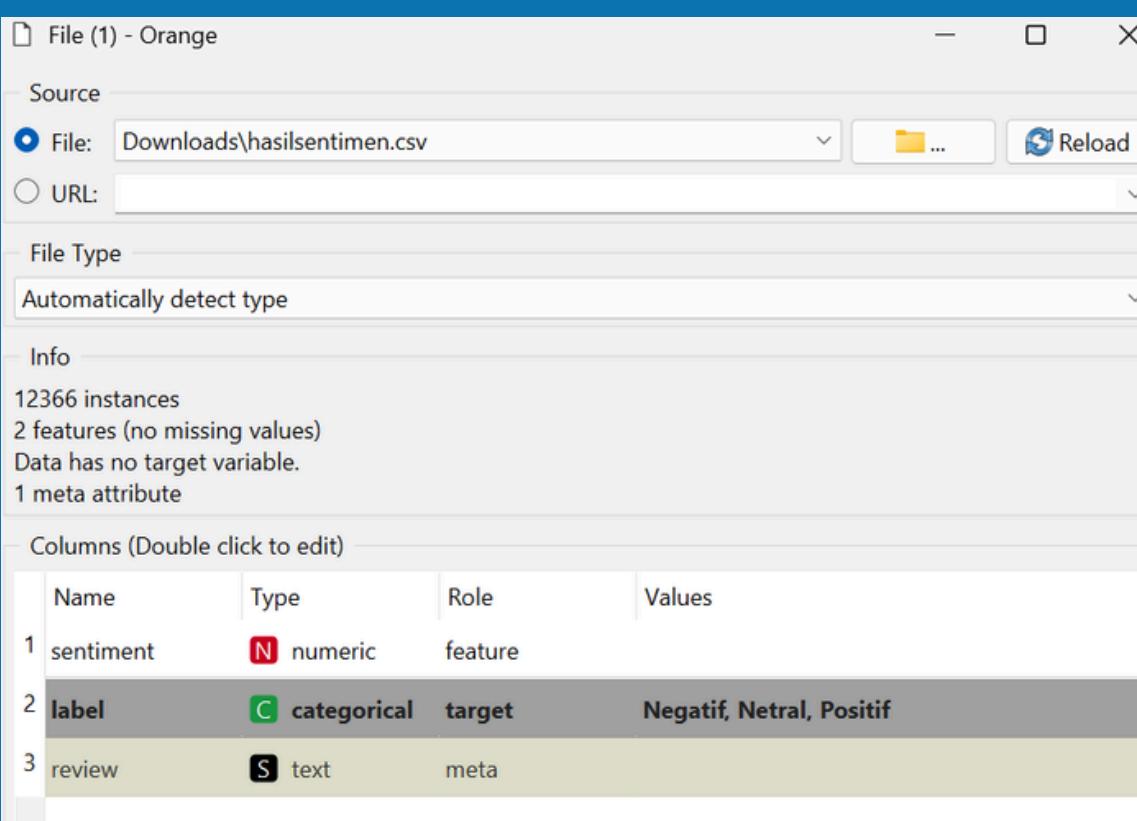
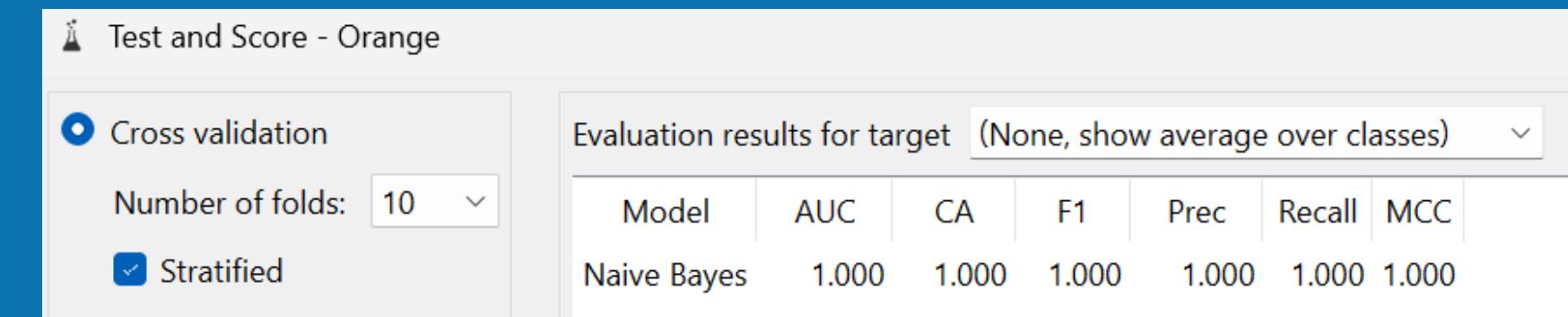
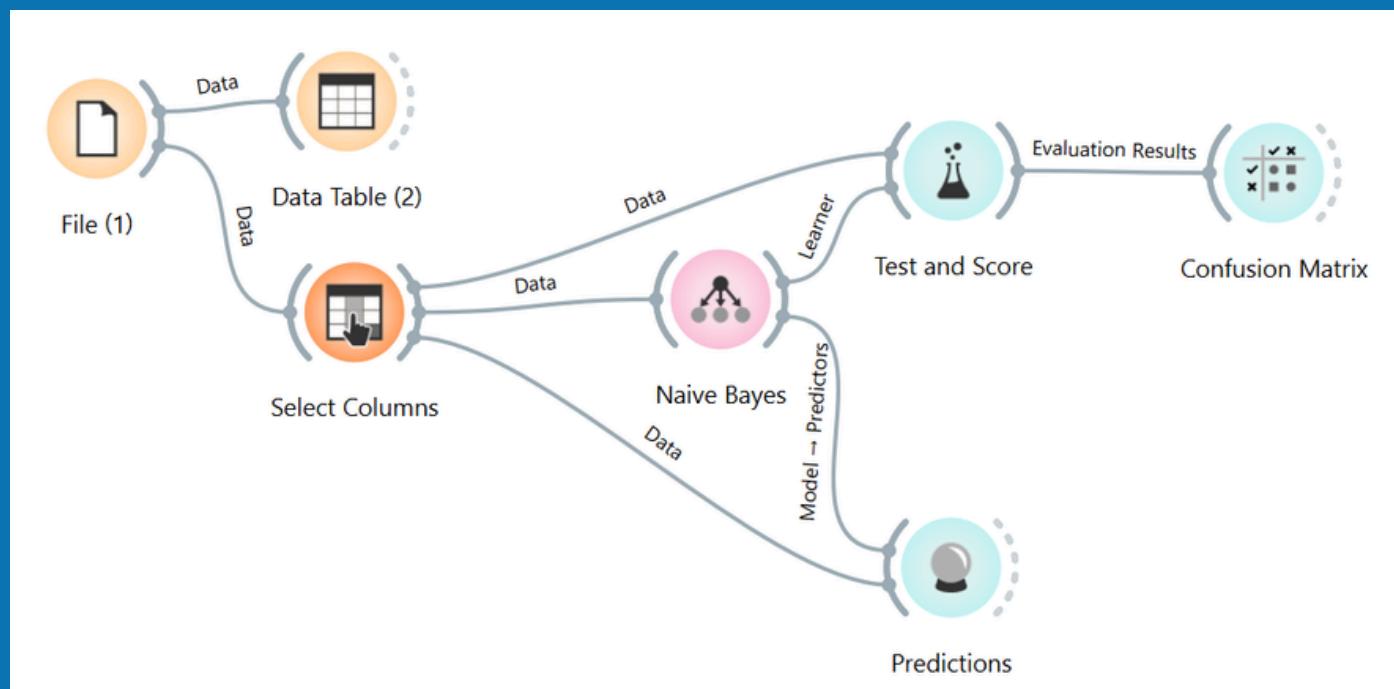
Proses ini menjadi **fondasi penting** untuk **membangun model machine learning** yang dapat **mengklasifikasikan sentimen** atau **menemukan pola** dalam data teks





- Widget **Tweet Profiler** digunakan untuk **menganalisis emosi** yang terkandung dalam ulasan pengguna **berdasarkan model Ekman (Multi Class)**, yang mengkategorikan emosi menjadi enam kelas utama, yaitu **Anger (Marah)**, **Disgust (Jijik)**, **Fear (Takut)**, **Joy (Senang)**, **Sadness (Sedih)**, dan **Surprise (Kaget)**
- Hasil analisis menunjukkan bahwa **majoritas** ulasan mengandung emosi **Joy (senang)** sebanyak 7.071 kemunculan, diikuti oleh **Surprise (kaget)** dan **Fear (takut)**
- Emosi **negatif** seperti **Anger**, **Disgust**, dan **Sadness** memiliki **jumlah yang lebih sedikit**, namun tetap signifikan untuk diperhatikan sebagai indikasi masalah yang dialami pengguna
- Widget **Box Plot** digunakan untuk **memvisualisasikan distribusi dan variasi emosi** dalam data ulasan secara keseluruhan, membantu **memahami rentang intensitas emosi pengguna** yang muncul

MODELLING WORKFLOW



- Pada tahap **modeling**, digunakan **algoritma Naive Bayes** untuk mengklasifikasikan **sentimen ulasan pengguna** aplikasi Livin' by Mandiri menjadi **tiga kelas: Negatif, Netral, dan Positif**
- **Data** yang digunakan **telah diproses** dan **fitur-fitur teks** telah **diekstraksi** sehingga siap untuk pelatihan model
- **Widget Select Columns** digunakan untuk **memilih atribut yang relevan sebagai fitur dan label target untuk klasifikasi**
- Evaluasi model dilakukan menggunakan **metode Cross Validation** dengan **10 folds** untuk menguji **keakuratan dan konsistensi model**

MODELLING WORKFLOW

Test and Score - Orange

Cross validation

Number of folds: 10

Stratified

Model	AUC	CA	F1	Prec	Recall	MCC
Naive Bayes	1.000	1.000	1.000	1.000	1.000	1.000

Hasil evaluasi menunjukkan **performa Naive Bayes yang sangat baik**, dengan skor :

- **Akurasi (CA)** : 1.000
- **F1-Score** : 1.000
- **Precision** : 1.000
- **Recall** : 1.000

Confusion Matrix menunjukkan bahwa semua data diklasifikasikan dengan benar **tanpa kesalahan**, yang mengindikasikan **model sangat efektif** pada dataset ini

		Predicted			Σ
		Negatif	Netral	Positif	
Actual	Negatif	3892	0	0	3892
	Netral	0	4885	0	4885
	Positif	0	0	3589	3589
		Σ	3892	4885	3589
					12366

MODELLING WORKFLOW

Predictions - Orange		
Show probabilities for Classes in data		
	Naive Bayes	error
11807	1.00 : 0.00 : 0.00 → Negatif	0.001
11808	1.00 : 0.00 : 0.00 → Negatif	0.001
11809	1.00 : 0.00 : 0.00 → Negatif	0.001
11810	1.00 : 0.00 : 0.00 → Negatif	0.001
11811	1.00 : 0.00 : 0.00 → Negatif	0.001
11812	1.00 : 0.00 : 0.00 → Negatif	0.001
11813	1.00 : 0.00 : 0.00 → Negatif	0.001
11814	1.00 : 0.00 : 0.00 → Negatif	0.001
11815	1.00 : 0.00 : 0.00 → Negatif	0.001
11816	1.00 : 0.00 : 0.00 → Negatif	0.001
11817	1.00 : 0.00 : 0.00 → Negatif	0.001
11818	1.00 : 0.00 : 0.00 → Negatif	0.001
11819	1.00 : 0.00 : 0.00 → Negatif	0.001
11820	1.00 : 0.00 : 0.00 → Negatif	0.001
11821	1.00 : 0.00 : 0.00 → Negatif	0.001
11822	1.00 : 0.00 : 0.00 → Negatif	0.001
11823	1.00 : 0.00 : 0.00 → Negatif	0.001
11824	1.00 : 0.00 : 0.00 → Negatif	0.001
11825	1.00 : 0.00 : 0.00 → Negatif	0.001
11826	1.00 : 0.00 : 0.00 → Negatif	0.001
11827	1.00 : 0.00 : 0.00 → Negatif	0.001
11828	1.00 : 0.00 : 0.00 → Negatif	0.001
11829	1.00 : 0.00 : 0.00 → Negatif	0.001

label review sentiment

Widget **Predictions** menampilkan hasil prediksi sentimen untuk masing-masing ulasan secara rinci, termasuk nilai probabilitas dan identifikasi kesalahan klasifikasi (jika ada)

Hasil ini memperkuat keandalan model dalam memahami sentimen ulasan pengguna aplikasi Livin' by Mandiri

Widget **Predictions** menampilkan hasil prediksi sentimen untuk masing-masing ulasan secara rinci, termasuk nilai probabilitas dan identifikasi kesalahan klasifikasi (jika ada)

Hasil ini memperkuat keandalan model dalam memahami sentimen ulasan pengguna aplikasi Livin' by Mandiri



KELOMPOK 2 - SISTEM TEMU KEMBALI INFORMASI

KEY FINDINGS

Distribusi Panjang Review

- Mayoritas ulasan memiliki panjang sekitar **5–20 kata**, dengan beberapa ulasan yang sangat panjang (**>60 kata**)
- Hal ini menunjukkan bahwa meskipun banyak pengguna memberikan feedback singkat, terdapat juga yang **memberikan komentar rinci dan detail**

Kata Kunci Dominan

- Kata-kata yang **paling sering muncul dan penting** berdasarkan TF-IDF adalah: **update, aplikasi, livin, transaksi, bagus, mandiri, transfer, mudah, mantap**
- Hal ini mengindikasikan bahwa pengguna banyak membahas **fitur aplikasi dan pengalaman transaksi** sebagai aspek utama ulasan



Sentimen Pengguna

- Analisis sentimen menunjukkan bahwa **sebagian besar ulasan berisi sentimen positif**, namun ada juga sentimen negatif dan netral yang signifikan
- Sentimen **positif** mencerminkan **kepuasan terhadap kemudahan penggunaan dan fitur**, sedangkan sentimen **negatif** terkait dengan **gangguan teknis seperti error dan bug**

Emosi Pengguna

- Berdasarkan **analisis emosi** dengan model **Ekman**, mayoritas ulasan mengandung emosi **joy (senang)**, diikuti oleh **surprise (kaget)** dan **fear (takut)**
- Emosi **negatif** seperti **marah, jijik, dan sedih** muncul lebih **sedikit**, namun tetap relevan untuk evaluasi perbaikan aplikasi

Topik Utama dalam Ulasan

- Model topic **modelling** mengidentifikasi **10 topik utama** yang meliputi pembahasan terkait **update aplikasi, saldo, proses transfer, fitur, keamanan, gangguan teknis, dan lain-lain**
- Topik-topik ini membantu **mengelompokkan fokus pengguna** dan memberikan arahan untuk **pengembangan produk**

Kinerja Model Klasifikasi

- Model **Naive Bayes** berhasil mengklasifikasikan sentimen dengan **akurasi sempurna (100%)**, menunjukkan bahwa fitur teks yang diekstraksi **sangat representatif** untuk **perbedaan kelas sentimen**
- **Confusion matrix** memperkuat bahwa **tidak ada prediksi yang salah** dalam dataset evaluasi





CONCLUSION

KESIMPULAN

- 1 **Mayoritas pengguna memberikan ulasan positif**, menunjukkan tingkat kepuasan yang cukup tinggi terhadap fitur dan kemudahan penggunaan aplikasi. Namun, ulasan negatif yang ada memberikan sinyal penting terkait **masalah teknis seperti gangguan sistem dan error yang harus diperbaiki**
- 2 Topik utama yang dibahas meliputi **pembaruan aplikasi, proses transaksi, dan fitur keamanan**, sehingga pengembang dapat memprioritaskan peningkatan pada area-area ini untuk mempertahankan dan meningkatkan kepuasan pengguna
- 3 **Emosi pengguna yang dominan adalah rasa senang dan terkejut**, yang mengindikasikan pengalaman **positif** dan **antusiasme** terhadap aplikasi. Namun, **emosi negatif** seperti **marah** dan **takut** menunjukkan adanya **frustrasi** yang perlu ditangani secara serius
- 4 Model klasifikasi **Naive Bayes** menunjukkan **performa sangat baik** dalam mengkategorikan sentimen, yang menguatkan **keakuratan analisis** dan **potensi penerapan model** ini untuk pemantauan ulasan secara otomatis di masa depan



IMPLIKASI

1

Tim pengembang aplikasi dapat menggunakan hasil analisis ini sebagai basis untuk perbaikan produk dan layanan, terutama fokus pada mengurangi bug dan meningkatkan stabilitas aplikasi

2

Monitoring sentimen dan topik secara berkala dapat membantu dalam mengantisipasi masalah dan memahami kebutuhan pengguna secara real-time

3

Analisis ini juga membuka peluang untuk mengoptimalkan komunikasi pemasaran dan pelayanan pelanggan berdasarkan tema dan sentimen yang ditemukan dalam ulasan



**THANK YOU
FOR YOUR
ATTENTION**

