

DATA SCIENCE PORTFOLIO

Telegram App Review Analytics

Comprehensive NLP analysis of 100K user reviews using BERTopic for topic modeling & Logistic Regression for sentiment classification

By: **Az-Zukhrufu Fi Silmi Suwondo**



Project Overview



Problem Statement

Telegram app experiences 60.1% negative reviews from 100K user feedback. Keywords "login", "code", "premium", "pay" indicate authentication and monetization issues requiring systematic analysis.



Goal

Develop an automated NLP system achieving $\geq 85\%$ F1-score for key sentiments (Negative & Positive) and identifying top 5 actionable topics per class from 100K Telegram reviews.



Objectives

- Classify sentiment into negative, neutral & positive
- Extract topics from negative and positive reviews
- Deliver actionable insights for product team

Data Processing



NLP & ML



Visualization

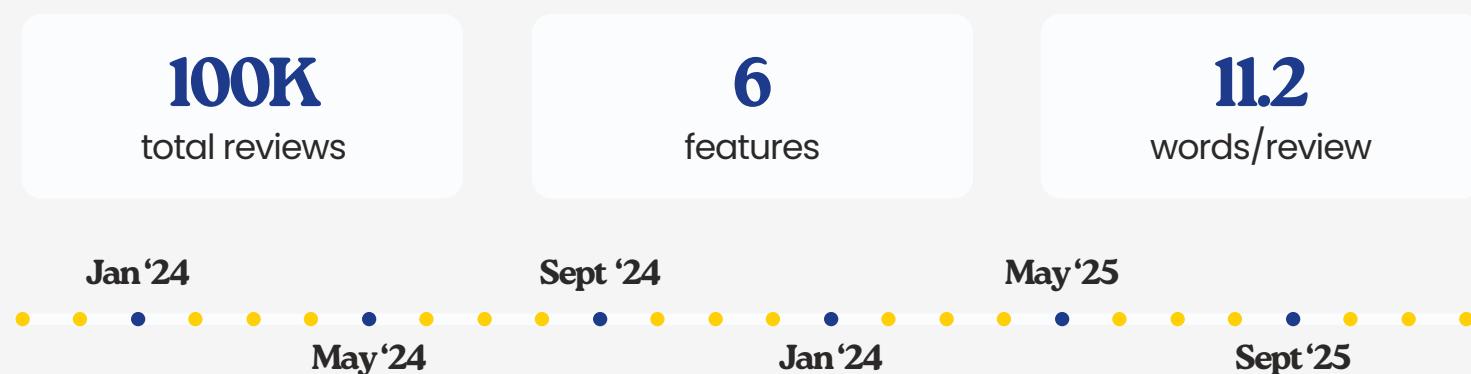


Deployment

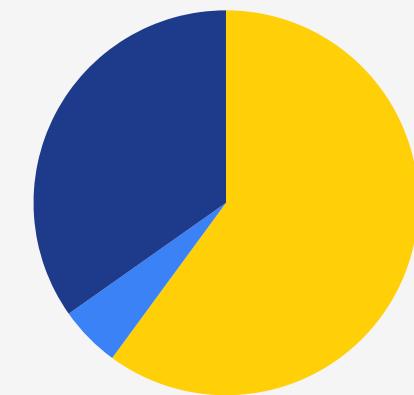


Dataset Overviews

Dataset Size



Target Distribution



Positive Ratings above 3	34.8%
Neutral Ratings equal to 3	5.2%
Negative Ratings below 3	60.1%

Data Structure

 Nama User	STRING
 Ulasan	TEXT
 Rating	INT

 Tanggal	DATETIME
 Likes	INT

 Versi App	STRING
--	--------

Data Structure

Completeness
Missing: Versi App (34%) **66.0%**

Text Issues
Reviews < 3 words **24.1%**

Duplicates
Text duplicates detected **19.9%**

Inconsistency
Rating-sentiment mismatch **5.9%**

Methodology Flow



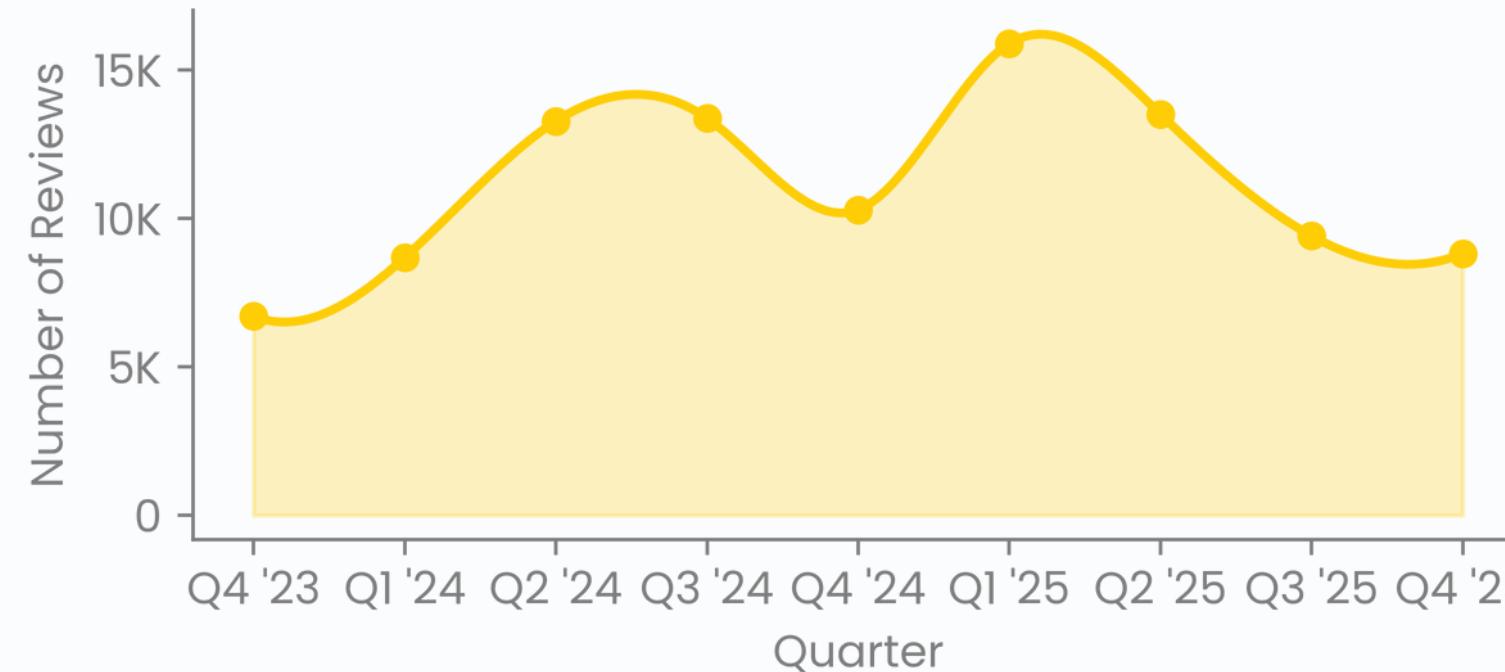
Sentiment Classification

- TF-IDF feature extraction with n-grams
- 10+ algorithms comparison
- Hyperparameter tuning with Optuna
- Threshold calibration & evaluation

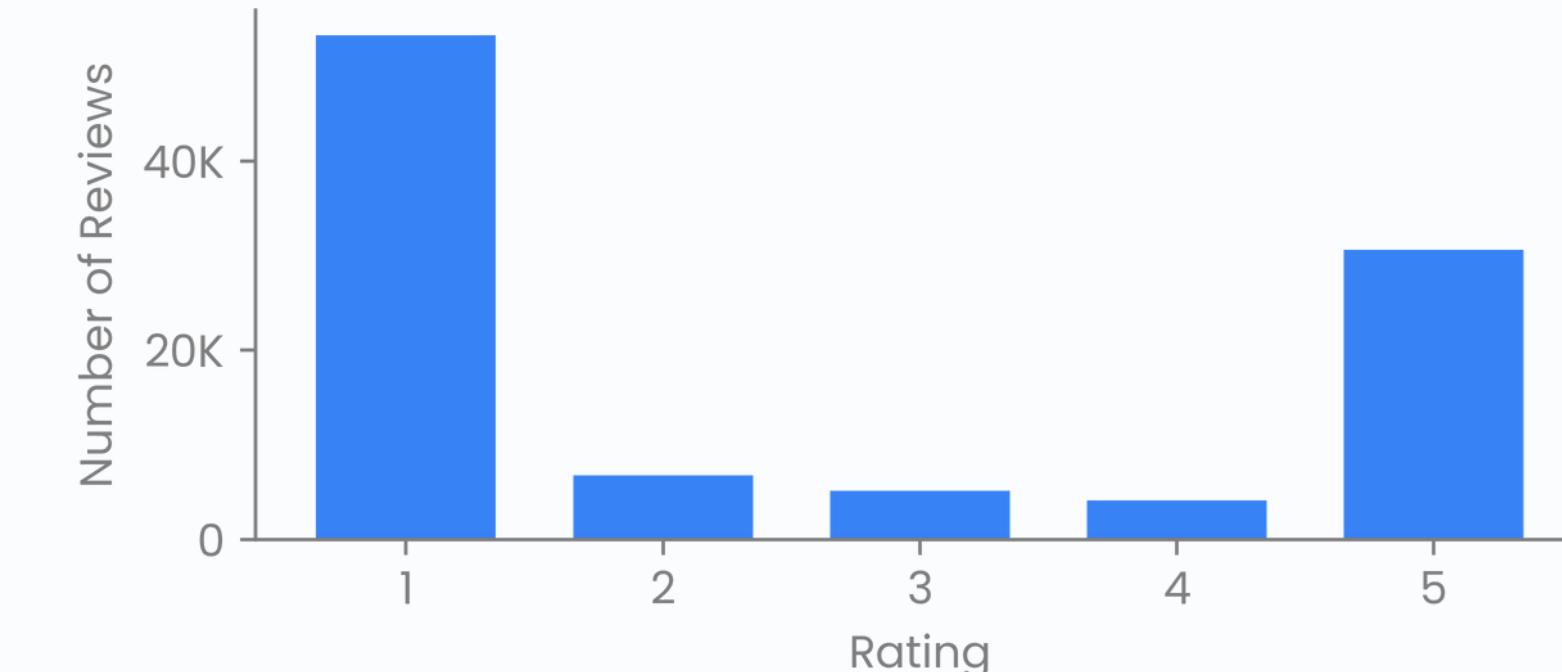
Topic Modeling

- Separate Positive & Negative reviews
- BERTopic with multilingual embeddings
- Coherence-based topic optimization
- Priority matrix & actionable insights

Polarized User Sentiment



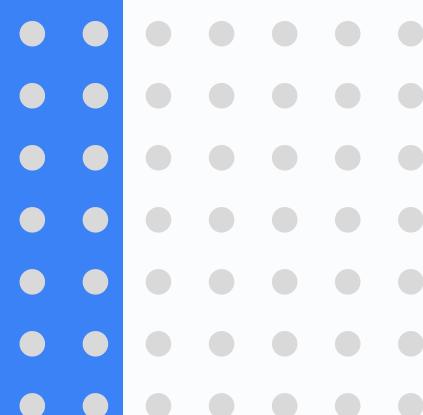
Traffic peaked in Q1 '25 (15.9k) but failed to sustain. Volume surged **+83% YoY in early 2025** before **dropping 44%** by year-end, signaling a **major influx** followed by significant **user churn**.



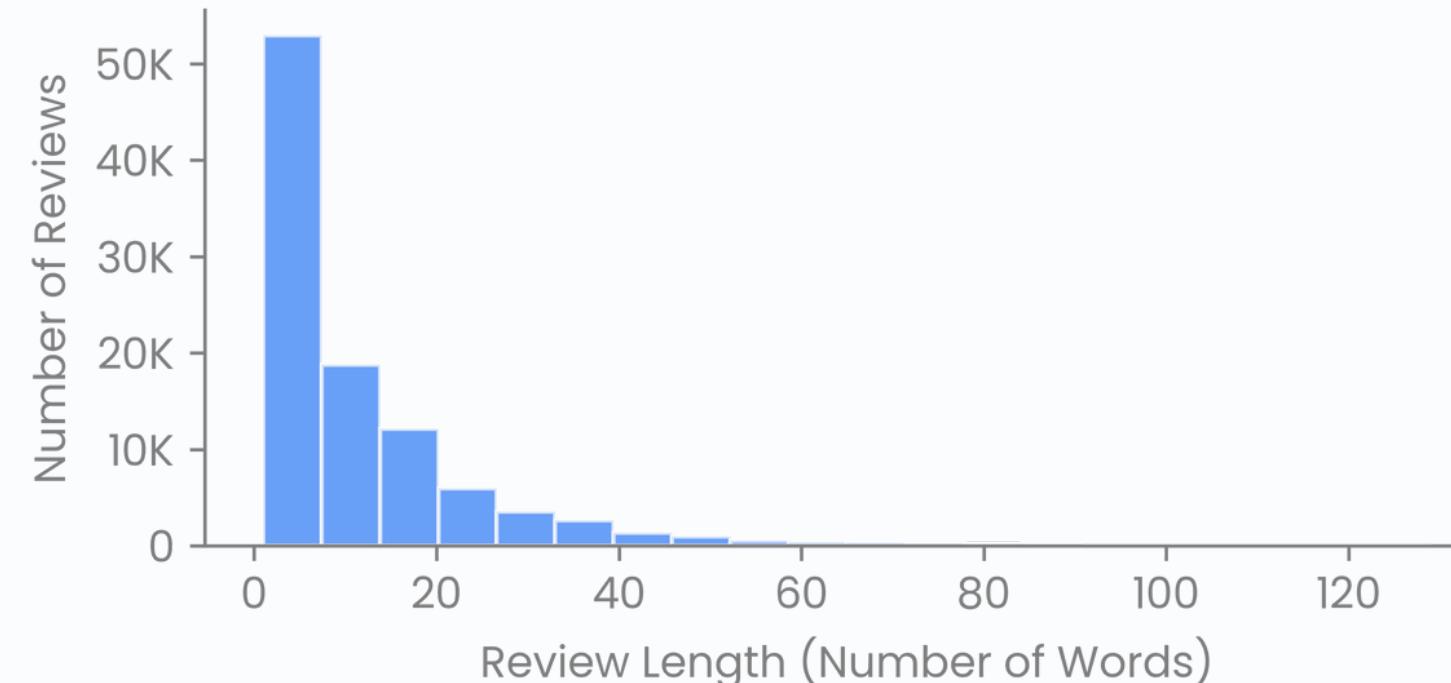
Critical Sentiment shows **53% of reviews are 1-Star**. Ratings are strictly **polarized in a U-shaped pattern**, where the overwhelming majority of negative feedback outweighs the 30% positive user base, indicating **severe product issues**.

💡 Recommendation

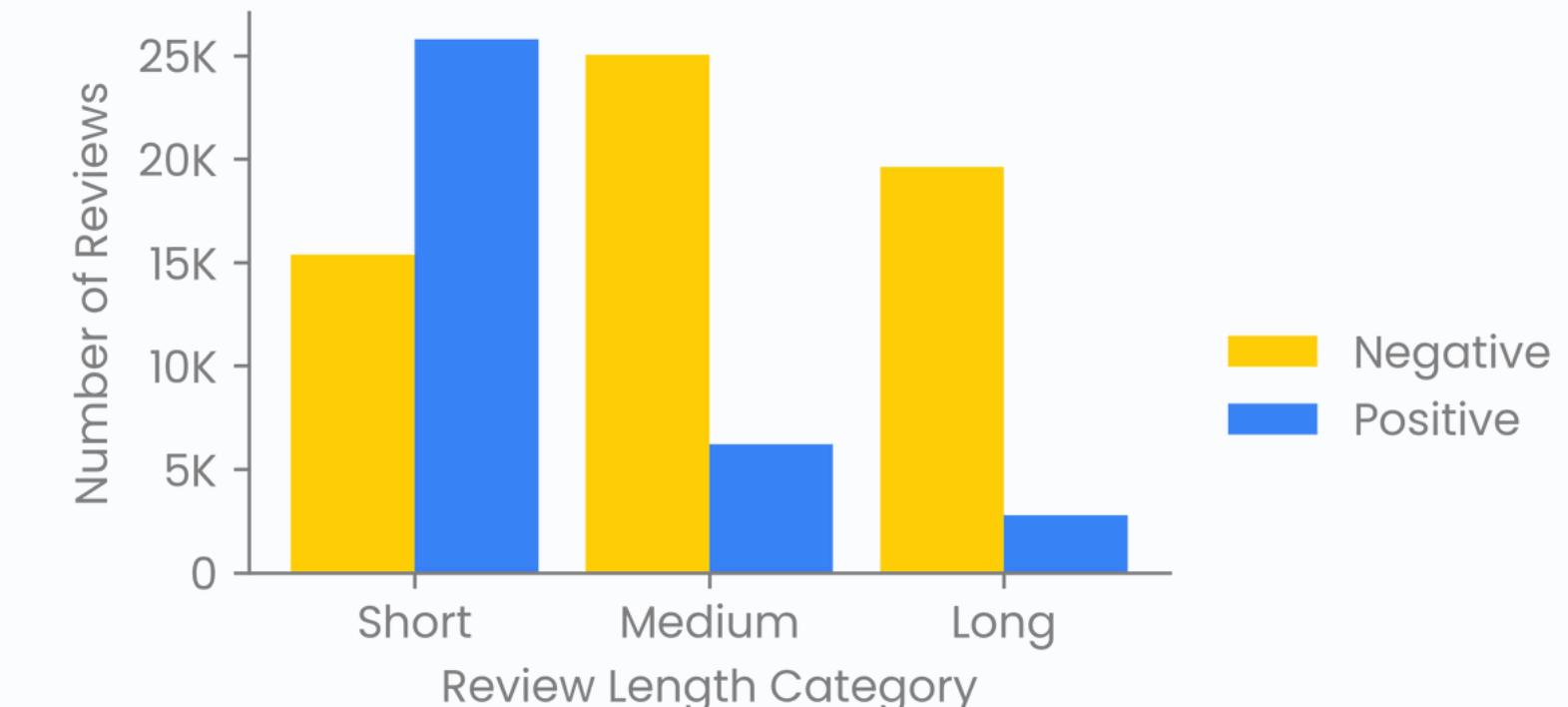
- Correlate Q1 '25 spike with app releases or marketing campaigns to identify trigger
- Conduct topic modeling on 53k 1-star reviews to pinpoint top complaints
- Analyze retention data to confirm if Q4 drop is driven by dissatisfied users



Length Signals Sentiment



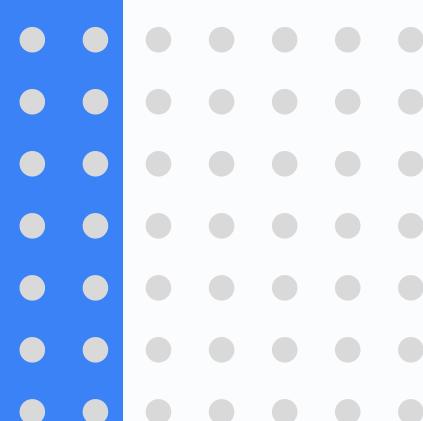
Unhappy users write **2.7x more**. **Negative reviews** average 14.4 words, significantly **longer than positive** ones (5.4 words). Users tend to elaborate extensively on complaints, while positive feedback is typically brief (e.g., "Good app").



Longer reviews are overwhelmingly **negative**. While positive sentiment leads in the "Short" category, negative reviews dominate the "Medium" and "Long" categories. If a user writes a long review, it is highly likely to be a **complaint**.

💡 Recommendation

- Target long negative reviews for NLP analysis to extract specific root causes
- Focus on long positive reviews for insights; ignore short ones as they lack detail
- Treat neutral reviews as complaints, since their length pattern mirrors negative feedback





Words Reveal Problems



Access blocked at the door. 7 out of the top 10 negative terms (kode, login, sms, otp, verifikasi) relate directly to **authentication failures**. Users are critically stuck in the login/signup loop and cannot access the app.



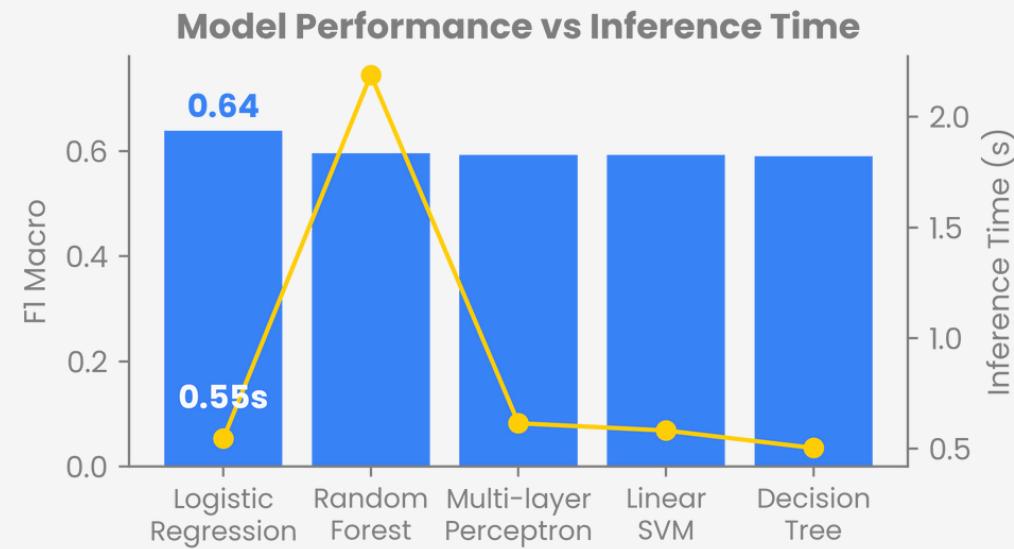
Satisfaction is high-level & generic. Feedback is dominated by **general praise** (bagus, good, mantap) rather than specific features, indicating users **rarely highlight distinct functional strengths**.



Recommendation

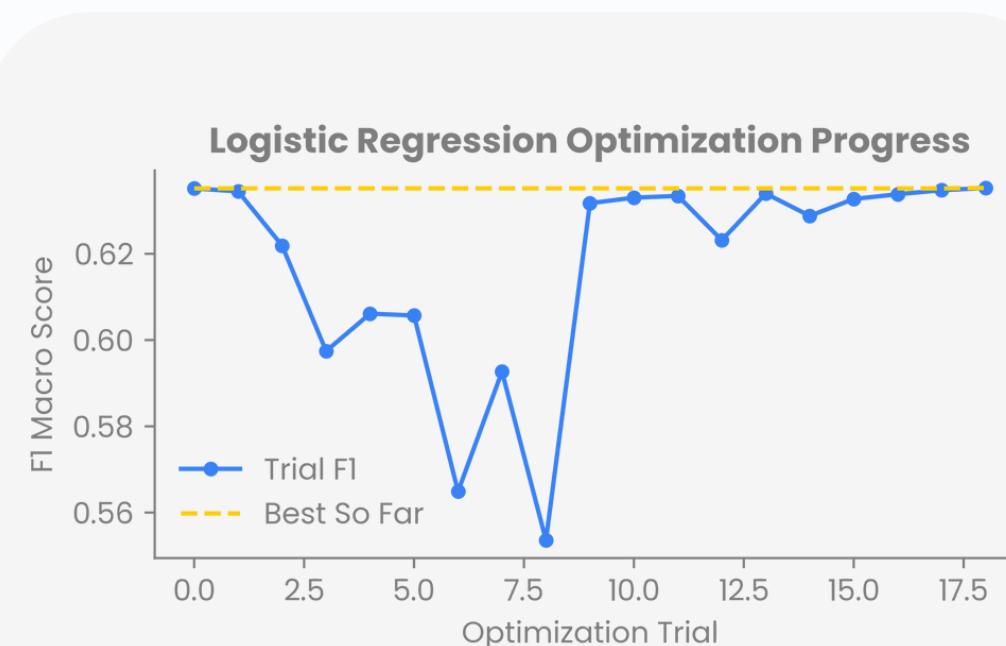
- Fix SMS/OTP infrastructure immediately to resolve verification blocks and restore user access
 - Audit "Premium" flows to prevent accidental user lockouts during subscription upgrade attempts
 - Prompt happy users for specific details to uncover product strengths for better roadmap prioritization

Classification Model Development



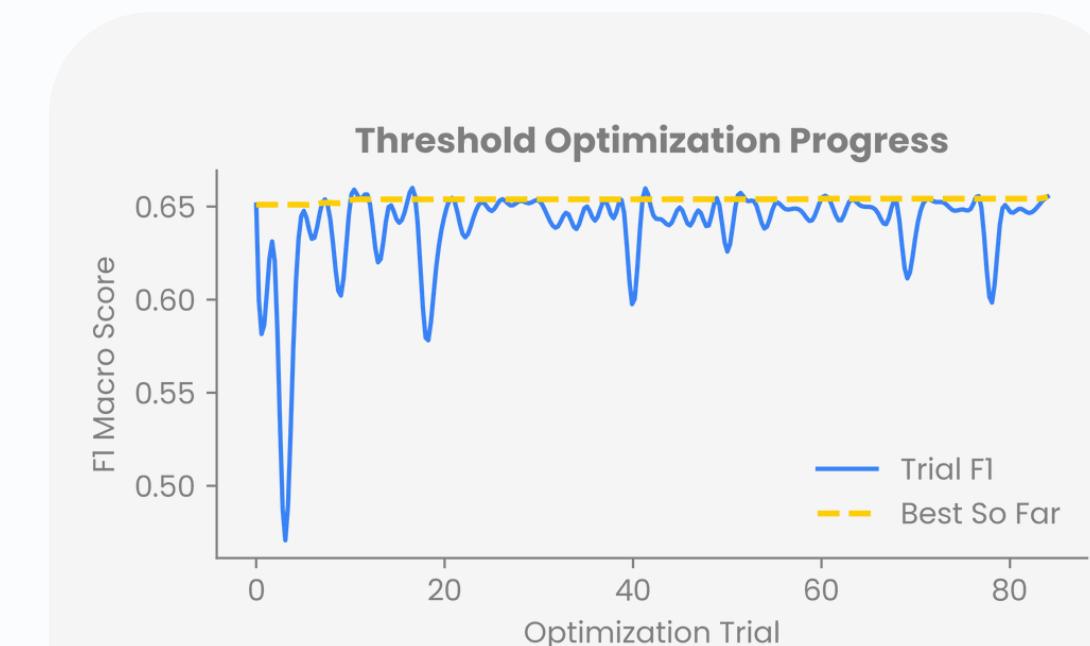
Model Selection

Evaluated **10 algorithms** via Stratified CV. **Logistic Regression** chosen for best F1-Macro (0.64) and speed (0.54s), outperforming complex models.



Hyperparameter Tuning

Employed **Optuna** for Bayesian optimization. Secured optimal regularization ($C=0.81$) and weights, stabilizing performance on sparse review data.



Threshold Calibration

Calibrated decision **thresholds per class**. Enhanced minority detection and boosted Accuracy from 80% to 85% (F1-Macro 0.66).

Model Validation Results

Performance Summary

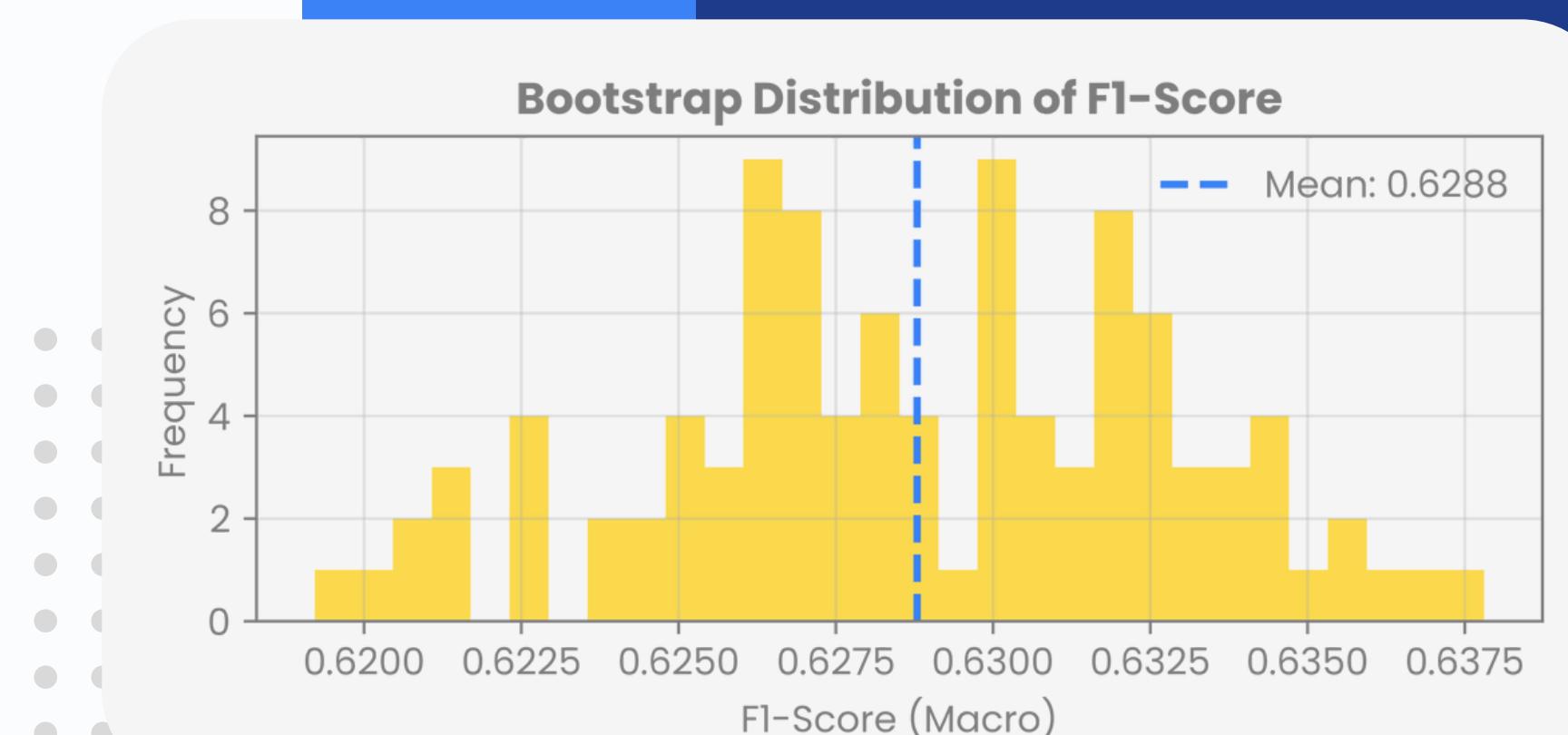
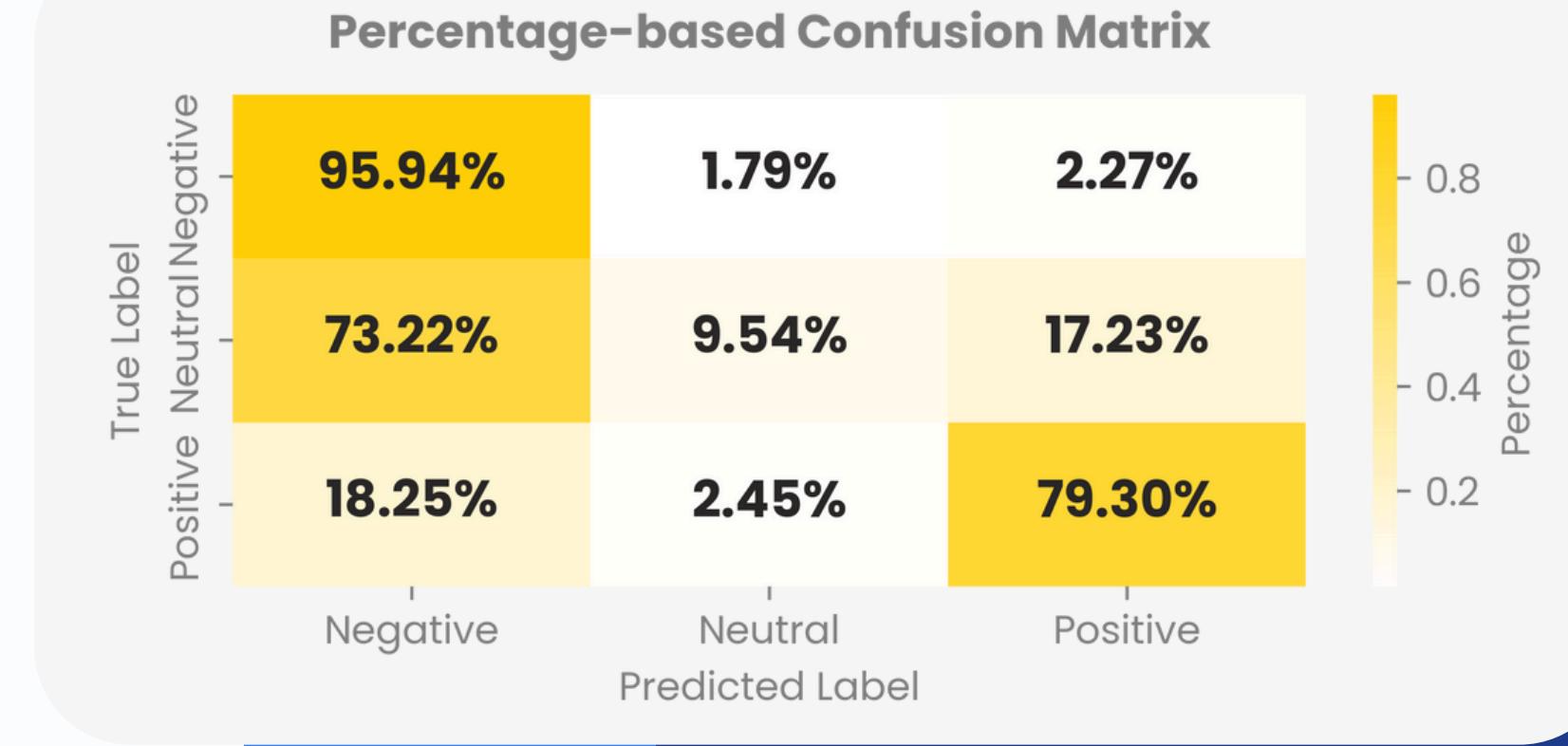
High reliability on priority classes. Achieved 96% recall for Negative and 79% for Positive reviews. Low overall F1 is solely due to the minority Neutral class (data scarcity).

Stability & Limitations

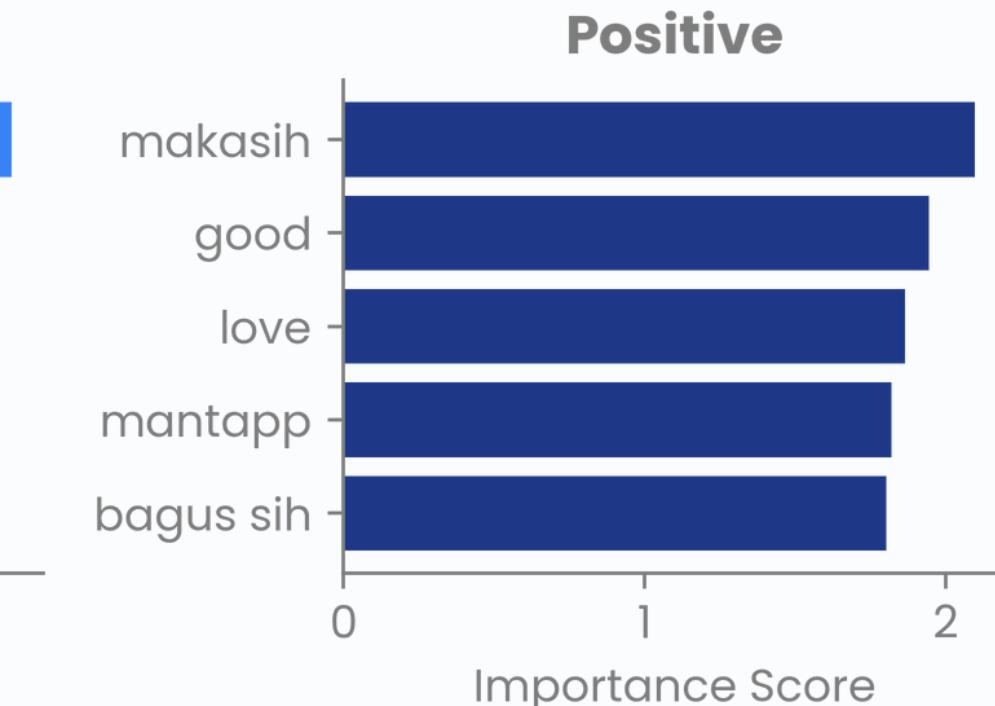
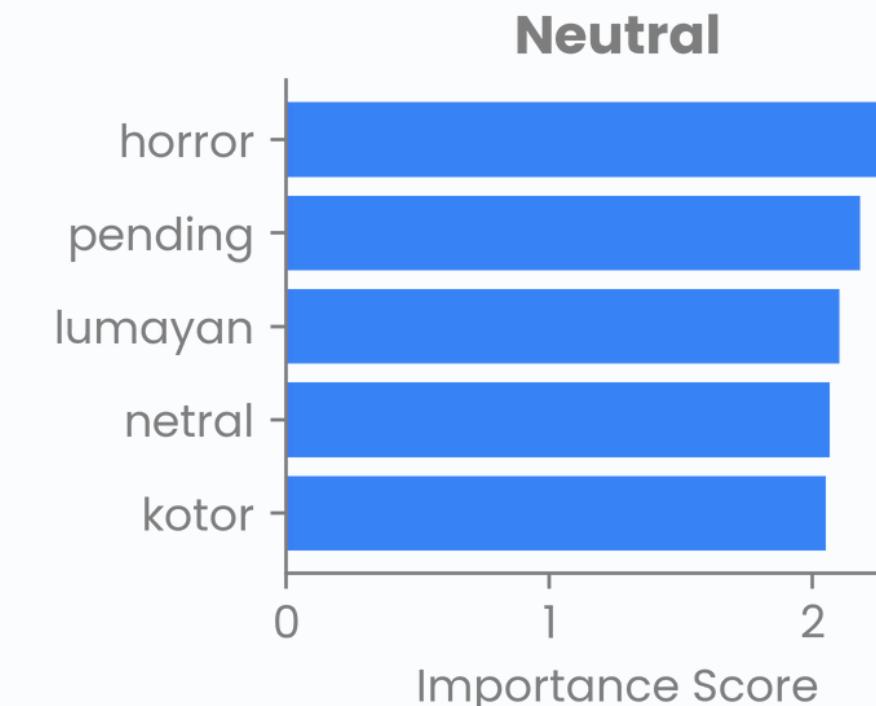
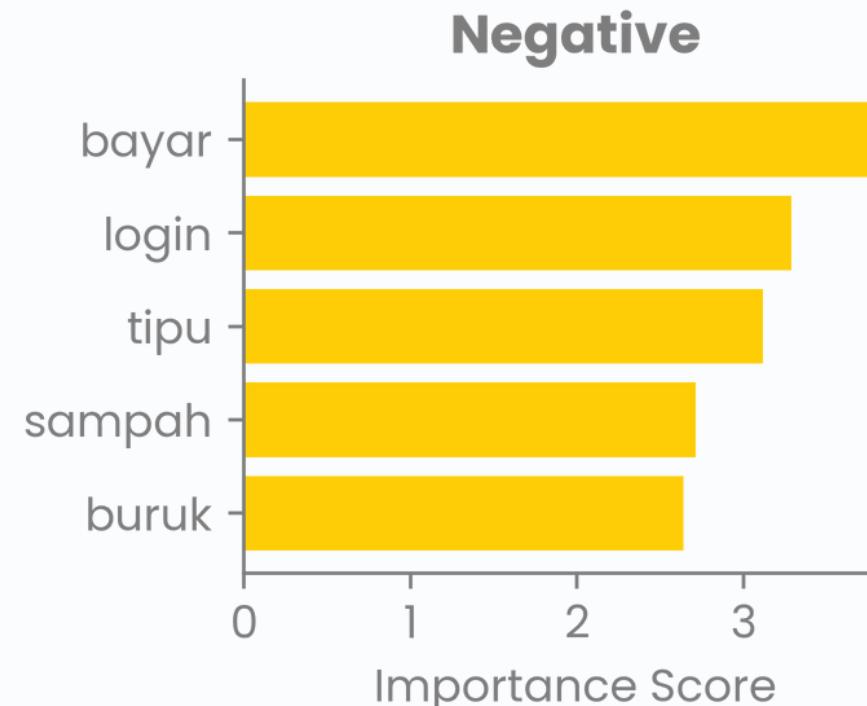
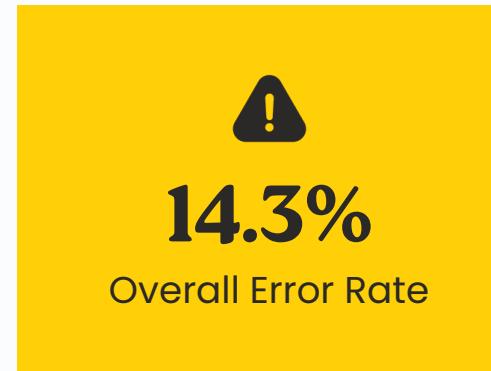
Statistically stable performance (CI: 0.62–0.63). Neutral class detection remains limited due to severe imbalance, but does not affect key sentiment analysis.

Next Steps

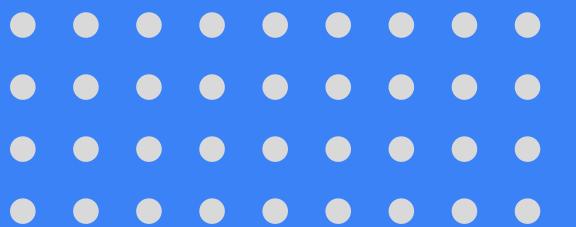
Focus topic extraction on Negative and Positive reviews to derive actionable insights. Future work: Acquire more labeled Neutral data.



Model Interpretation & Analysis



The model identifies negative sentiment through words like **bayar**, **login**, and **kode**, and positive sentiment with **mantap**, **bagus**, and **cepat**. It **misclassifies 14.25%** of reviews, notably 1,243 positive reviews with technical issues as negative. Neutral reviews are often classified as Negative (73%), acting as a fail-safe mechanism to ensure no potential user issues are overlooked. Despite **0% low-confidence predictions**, the model is reliable only for clear sentiments and requires more neutral data for mixed feedback.



Topic Modeling with BERTopic



Data Preparation

Split reviews by sentiment, remove noise, and filter using 500+ Indonesian stopwords

Embedding Generation

Convert reviews into 384-dimensional semantic vectors using multilingual-MinILM-L12-v2 model

Topic Discovery

HDBSCAN clusters similar reviews, then c-TF-IDF extracts top keywords with n-gram (1,2)

Quality Refinement

Grid search evaluates 5-20 topics using Coherence Score to find optimal count and reduce outliers

Validation & Insights

Priority matrix ranks topics by severity × frequency with diversity score validation

 **60K+**

Negative Reviews

 **34K+**

Positive Reviews

Objective & Impact

Extract specific themes from 60K negative and 35K positive Telegram reviews to transform **vague sentiment scores into actionable insights**. **Identify concrete product issues** like "OTP failures block 15% of users" and strengths like "Group chat drives 40% of positive feedback" for **data-driven roadmap prioritization**.

Critical Issues Breakdown

Topic Overview

Authentication & OTP Failure

kode sms verifikasi kode verifikasi otp

42.7%

Account Bans & System Errors

banned blokir eror internal kesalahan

8.7%

Monetization & Premium Issues

login bayar story premium kaya login premium

6.2%

Media Upload & Story Bugs

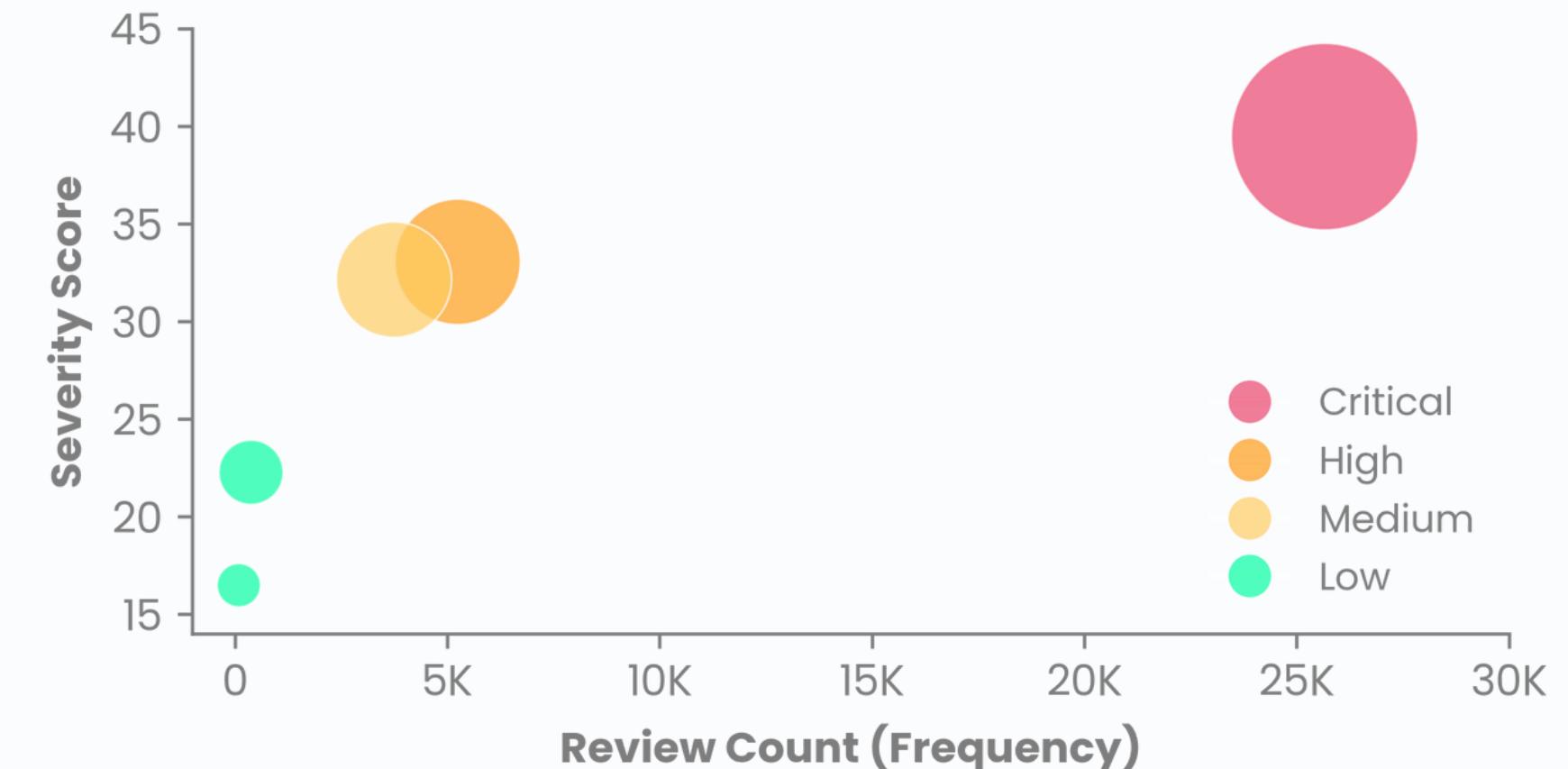
foto cerita profil status bug

0.6%

Translation & Spam Content

bahasa terjemahan translate online

0.1%



BERTopic analysis identified **five distinct topics** from **59,956 negative reviews** with **97.9% coverage**. Authentication **failures dominate at 42.7% severity**, requiring **immediate OTP infrastructure fixes** to restore access for 25K+ blocked users.

User Satisfaction Breakdown

Topic Overview

Premium Experience & Business Use

cepat prem informasi menyenangkan love

62.4%

Media Sharing & Groups

video file chat pengguna grup

8.5%

Account Support & Access Issues

blokir whatsapp log banned

8.2%

Entertainment, Gaming & Localization

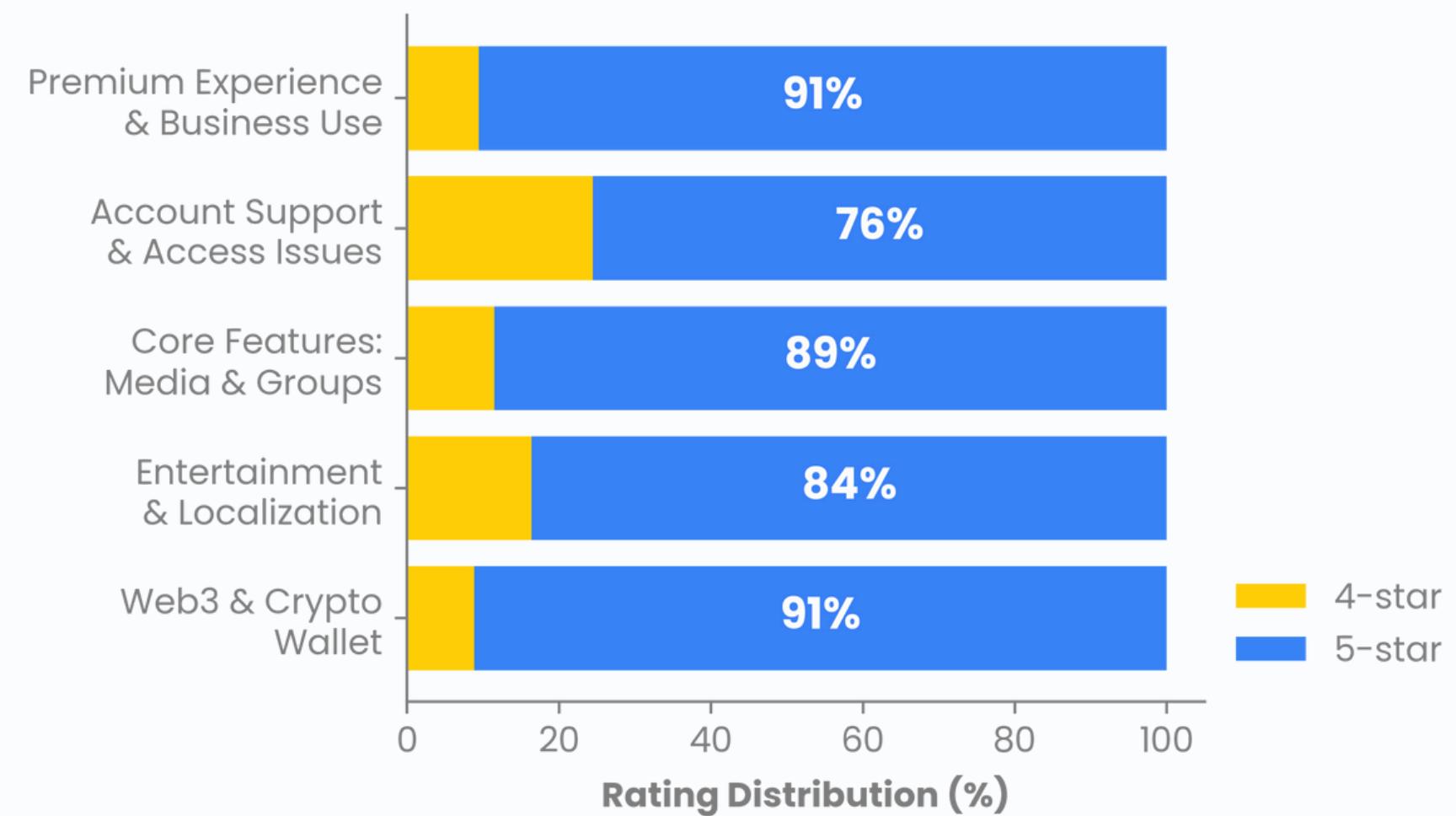
indonesia drama bahasa korea

0.2%

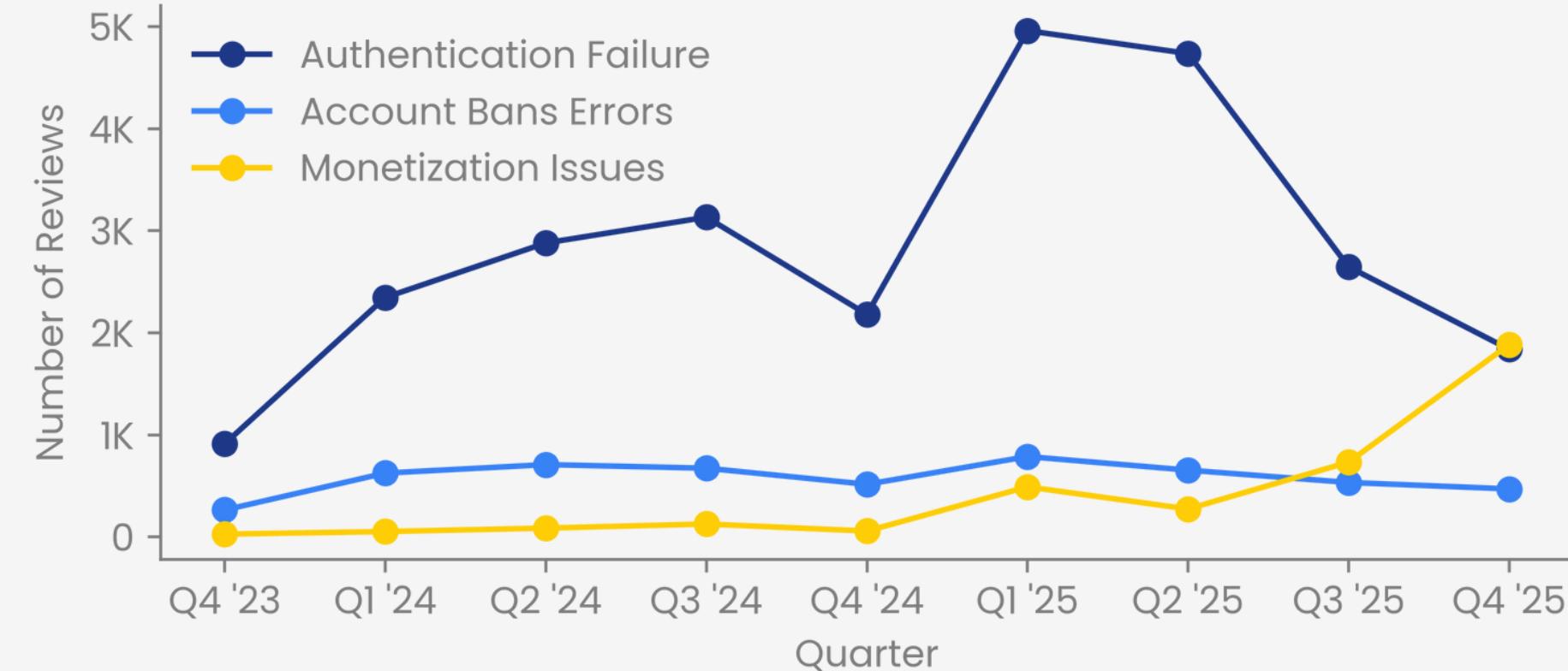
Web3, Crypto Wallet & Earning

wallet coin kerja usdt koin

0.04%



BERTopic analysis of **33,780 positive reviews** reveals **premium features dominate satisfaction** at 62.4% (21,069 reviews, rating 4.91). Core media sharing and group functionalities drive an additional 8.5% positive sentiment, validating **product-market fit**.



Topic Model Performance & Insights

Both models achieved **robust performance** with **over 95% coverage** and **strong topic separation**. While **Q1 2025** marked a **platform-wide crisis**, the **2.6x monetization complaint surge** in **Q4 2025** (733 to 1,885) is the most urgent issue requiring immediate policy investigation.

Strategic Recommendations



Fix Authentication

Source: 42.7% Negative Reviews

Audit SMS Gateway

Switch providers for regions with high OTP failure rates.

Alternative Login

Implement WhatsApp or Email verification to bypass SMS blockers.

Success Metric: OTP Delivery Rate



Stabilize Monetization

Alert: +2.6x Complaints in Q4

Investigate Q4 Updates

Review pricing changes or bugs released in Oct-Dec '25.

Prevent Lockouts

Fix "Login Premium" loops where paid users lose access.

Success Metric: Payment Ticket Vol.



Leverage "Premium"

Strength: 91% Satisfaction

Marketing Hook

Use "File Sharing" & "Groups" testimonials in ads.

B2B Focus

Prioritize features for Business users (highest retention segment).

Success Metric: Upgrade Conversion



Business Impact Summary

Recover Traffic

Resolving OTP blocks restores access for 25K+ stuck users.

Revenue Defense

Stopping Q4 monetization complaints protects key revenue streams.

Product Fit

Capitalizing on Core Features validates unique selling points.



Thank You

Explore my interactive tool for analyzing sentiment in Telegram app reviews with real-time classification and visualization

[Try the App](#)



afslimis@gmail.com



linkedin.com/in/az-zukhrufu-fi-silmi-suwondo/



github.com/afslimis/