

Identifikasi salah sasaran Penerima Beasiswa KIPK melalui cuitan X

Disusun oleh : Kelompok 6

Anggota Kelompok :

- Rizky Yanuar K (017)
- Dian Ayu Fauziah (011)

Latar Belakang

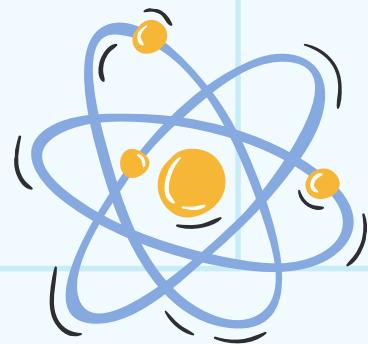


Dengan kemajuan teknologi saat ini, pendaftaran untuk mendapatkan beasiswa KIPK telah disederhanakan melalui situs web resmi Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. Hal ini memudahkan para calon pendaftar untuk mengakses informasi dan proses pendaftaran. Namun, disayangkan bahwa beberapa individu calon mahasiswa baru memanfaatkan kecanggihan dan kemudahan ini tanpa tanggung jawab. Mereka memanipulasi data dengan tujuan agar terlihat layak menerima beasiswa KIPK, meskipun dalam kehidupan nyata mereka memiliki gaya hidup hedonis dan tidak memenuhi syarat sebagai mahasiswa yang kurang mampu secara finansial.

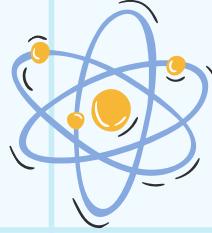
Tujuan

Proyek data mining ini bertujuan untuk mendeteksi kesalahan penargetan penerima beasiswa KIPK, dan hasilnya akan digunakan untuk memberikan wawasan tentang bagaimana program ini diterima oleh masyarakat

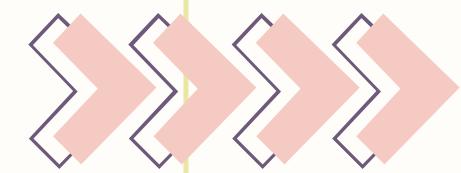
Manfaat



Inisiatif ini diarahkan untuk meningkatkan kejelasan dan tanggung jawab dalam alokasi beasiswa KIPK, melalui pemanfaatan analisis data yang netral dan faktual, yang menjamin proses pemilihan yang dilakukan dengan penuh integritas dan kesetaraan.



METODE



LSTM

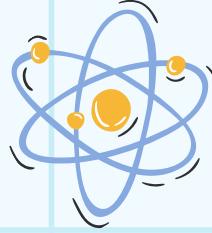
Metode Long Short-Term Memory (LSTM) digunakan karena memiliki beberapa keunggulan, antara lain LSTM mendapatkan akurasi yang baik saat pemrosesan data teks dan LSTM mampu memproses data yang relatif panjang.

Penjelasan Metode



Metode Long Short Term Memory (LSTM) merupakan salah satu metode deep learning yang banyak diaplikasikan untuk melihat sentimen pada data .

LSTM berkerja dengan membangun model yang terdiri dari beberapa lapisan, Masing-masing lapisan mengambil input dari yang lapisan sebelumnya dan memberikan output kepada lapisan berikutnya. Lapisan pertama mengambil urutan numerik sebagai input, dan lapisan terakhir memberikan label prediksi sebagai output.



FastText

metode pemodelan kata yang dikembangkan oleh Facebook AI Research. Pengembangan dari Word2Vec dengan memperhitungkan representasi kata sebagai kumpulan sub-kata (n-grams). Hal ini memungkinkan FastText untuk mengatasi kata-kata yang tidak dikenal dan bahasa yang jarang digunakan. Dalam analisis sentimen Twitter, FastText dapat digunakan untuk memahami makna teks dengan memperhitungkan struktur dan konteks sub-kata.

Perbedaan Word Embedding

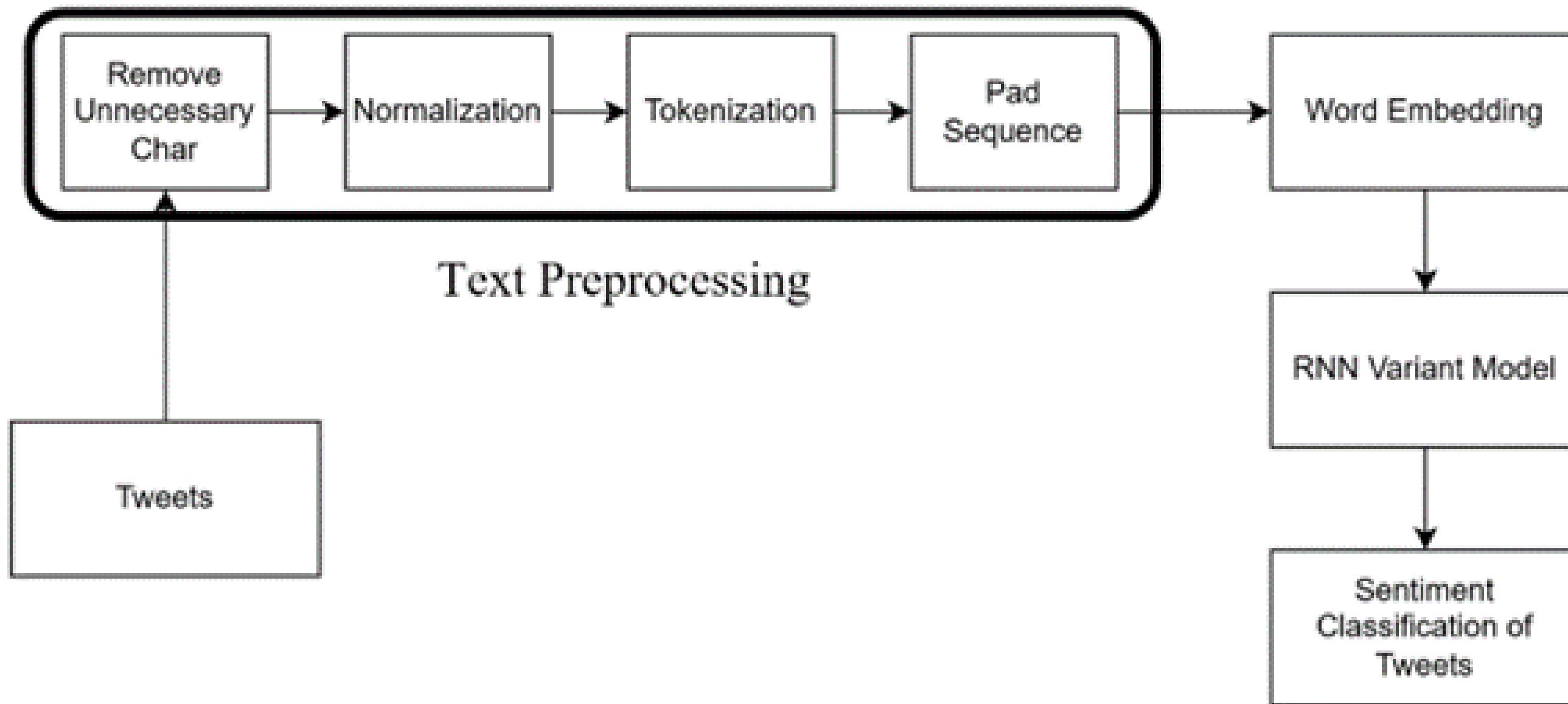
GloVe

metode yang mempelajari representasi vektor kata dengan memanfaatkan statistik ko-occurrence dari kata-kata dalam korpus teks. Ini mencoba untuk mengekstrak makna kata dari hubungan distribusi statistik mereka dalam konteks teks. Dengan GloVe, representasi vektor kata disesuaikan untuk merefleksikan makna dan hubungan antar kata dalam teks. Dalam analisis sentimen Twitter, GloVe dapat membantu dalam pemahaman konteks kata dan makna terkaitnya.

Dataset

	A	B	C	
1	conversation_id	created_at	favorite_count	full_text
2	1.79E+18	Tue Apr 30 2024 10:45:45.000 +0000 UTC	0	selain kasus salah sasaran kipk ada jd yg disebut disni yaitu pkh (gatau kepanjangannya apa) intinya selain uang dpt bahan' lain kek buah b
3	1.79E+18	Tue Apr 30 2024 10:45:45.000 +0000 UTC	0	Makin banyak yang ke spill yang kipk tapi ternyata hidupnya hedon alias salah sasaran. Kasian banget yang haknya direbut sama manusia-
4	1.79E+18	Tue Apr 30 2024 10:45:45.000 +0000 UTC	0	@personality404_the thing issss apakah that influencer mau diwawancara ttg kipk salah sasaran kek bjirr keknya orang biasa juga belum te
5	1.79E+18	Tue Apr 30 2024 10:45:45.000 +0000 UTC	0	Banyak amat kipk salah sasaran inimah yg salah jd jd yg seleksi kipk nya gasih
6	1.79E+18	Tue Apr 30 2024 10:45:45.000 +0000 UTC	0	@boomboomna @undipmenfess ini undip bnyk benar ee kejadian kipk salah sasaran
7	1.79E+18	Tue Apr 30 2024 10:45:45.000 +0000 UTC	0	Lagi rame masalah kipk yg salah sasaran ya? Pgn ikutan marah tapi energi udh terkuras abis sih sebenarnya. Wkwkwk. Cuman kuat berbag
8	1.79E+18	Tue Apr 30 2024 10:45:45.000 +0000 UTC	0	Beberapa hari ini banyak banget berita tentang kipk yang salah sasaran. Yang dapet orang-orang kaya sedangkan yang bener-bener berhak
9	1.79E+18	Tue Apr 30 2024 10:45:45.000 +0000 UTC	3	infokan link resmi pengaduan kipk salah sasaran di UNY... daripada spall spill identitas yg masih ngalor ngidul gajelas datanya
10	1.79E+18	Tue Apr 30 2024 10:45:45.000 +0000 UTC	0	Tp sebenarnya bagus krn banyak bgt kipk yg salah sasaran
11	1.79E+18	Tue Apr 30 2024 10:45:45.000 +0000 UTC	0	ga kelar' kejadian salah sasaran kipk ini klo di tuntas semua
12	1.79E+18	Tue Apr 30 2024 10:45:45.000 +0000 UTC	0	@tanyakanrl Mereka ngendorse satu produk berapa juta tuh dapat duitnya. Semoga sistem kipk diperketat lagi mayoritas penerima sal
13	1.79E+18	Tue Apr 30 2024 10:45:45.000 +0000 UTC	0	setiap lewat kipk yang salah sasaran gini gue sedih banget seriusan deh soalnya temen gue sampe berhenti kuliah karna gk dapat bdm wak

Alur Pengerjaan



Referensi



Neural Computing and Applications (2023) 35:9567–9578
<https://doi.org/10.1007/s00521-022-08186-1>

ORIGINAL ARTICLE



Emotion classification of Indonesian Tweets using Bidirectional LSTM

Aaron Glenn¹ · Phillip LaCasse¹ · Bruce Cox¹

Received: 24 May 2022 / Accepted: 22 November 2022 / Published online: 6 February 2023
This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023

Abstract

Emotion classification can be a powerful tool to derive narratives from social media data. Traditional machine learning models that perform emotion classification on Indonesian Twitter data exist but rely on closed-source features. Recurrent neural networks can meet or exceed the performance of state-of-the-art traditional machine learning techniques using exclusively open-source data and models. Specifically, these results show that recurrent neural network variants can produce more than an 8% gain in accuracy in comparison with logistic regression and SVM techniques and a 15% gain over random forest when using FastText embeddings. This research found a statistical significance in the performance of a single-layer bidirectional long short-term memory model over a two-layer stacked bidirectional long short-term memory model. This research also found that a single-layer bidirectional long short-term memory recurrent neural network met the performance of a state-of-the-art logistic regression model with supplemental closed-source features from a study by Saputri et al. [8] when classifying the emotion of Indonesian tweets.

Keywords DNN · BiLSTM · LSTM · Emotion classification · Machine learning

Dataset Primer



https://github.com/meisaputri21/Indonesian-Twitter-Emotion-Dataset/blob/master/Twitter_Emotion_Dataset.csv

Dataset Sekunder

<https://www.kaggle.com/datasets/anggapurnama/twitter-dataset-ppkm>

Accuracy model



Model

Generic

Glove

FastText

LSTM

0.768

0.777

0.782

GRU

0.761

0.766

0.78

biLSTM

-

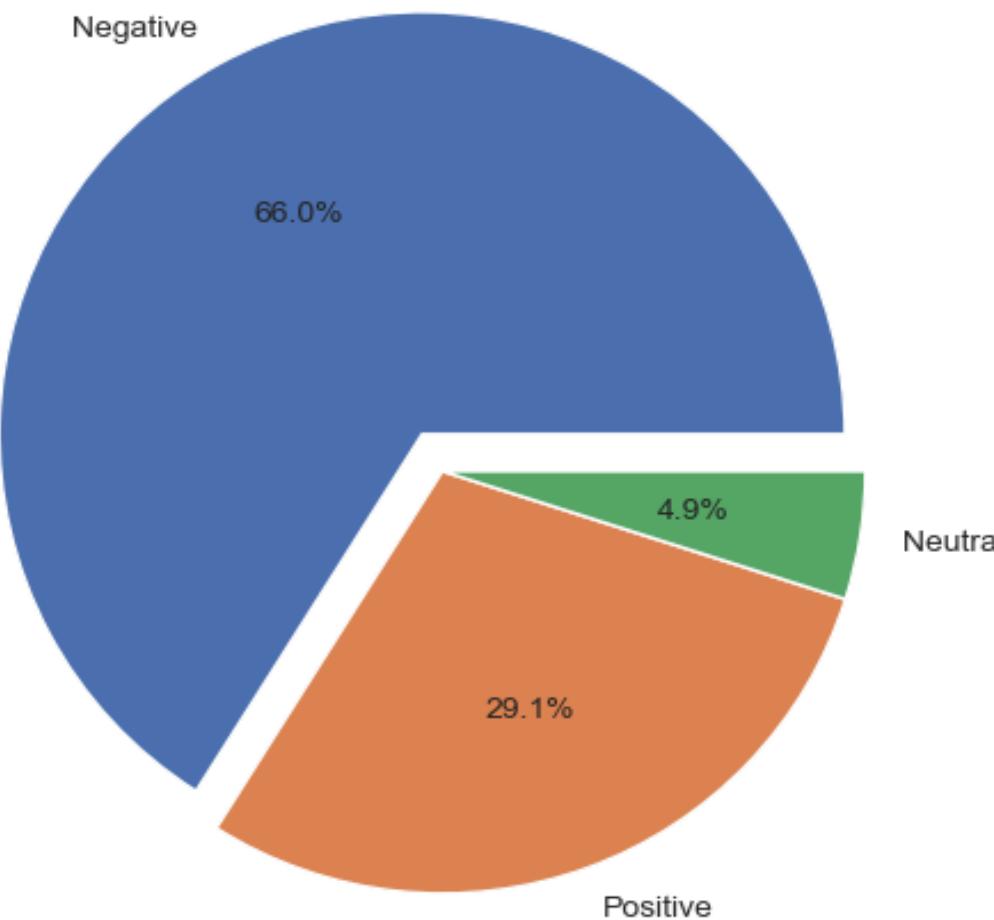
0.795

0.778

Hasil Distribusi sentimen

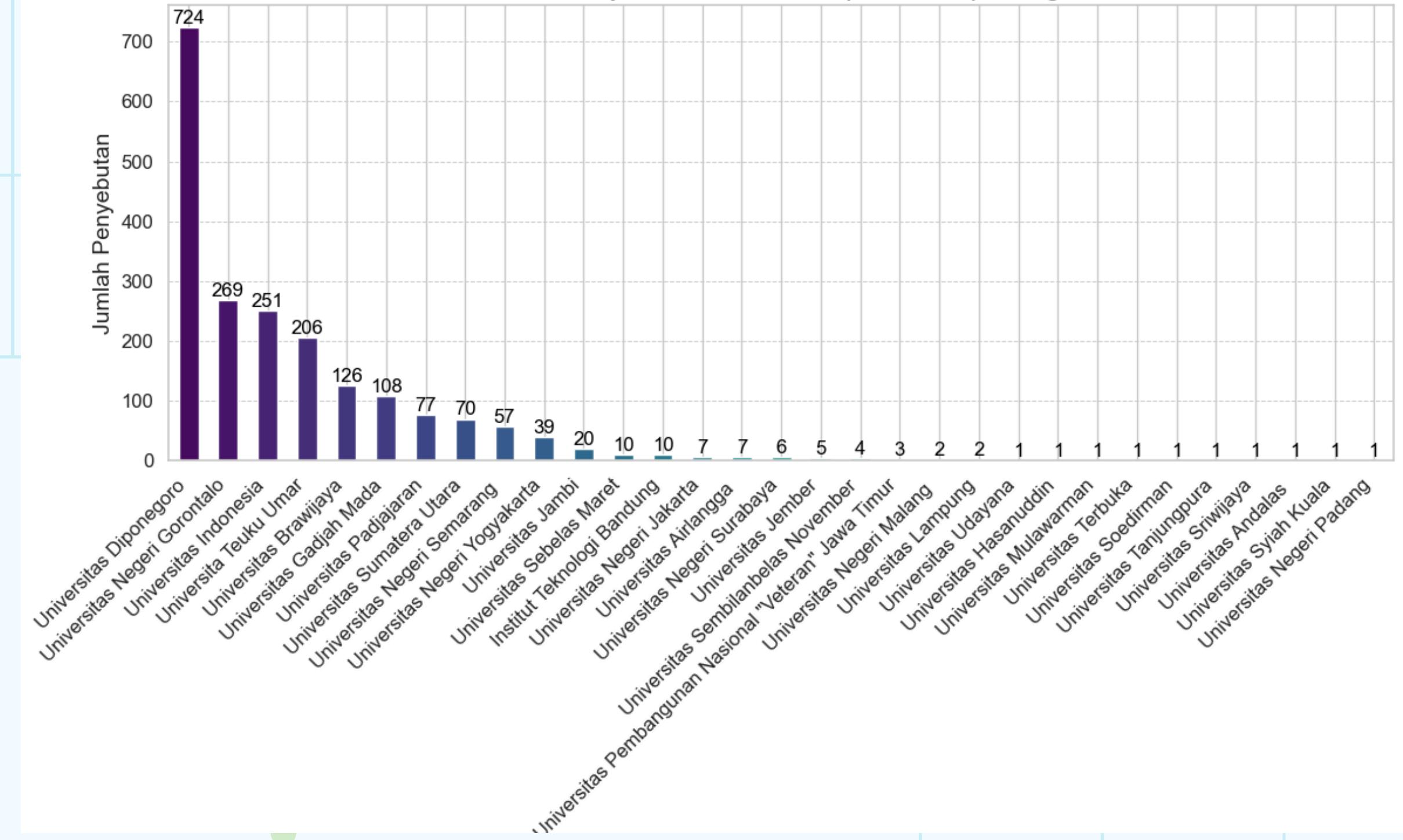


Sentiment Polarity on Twitter
(total = 3938 comments)



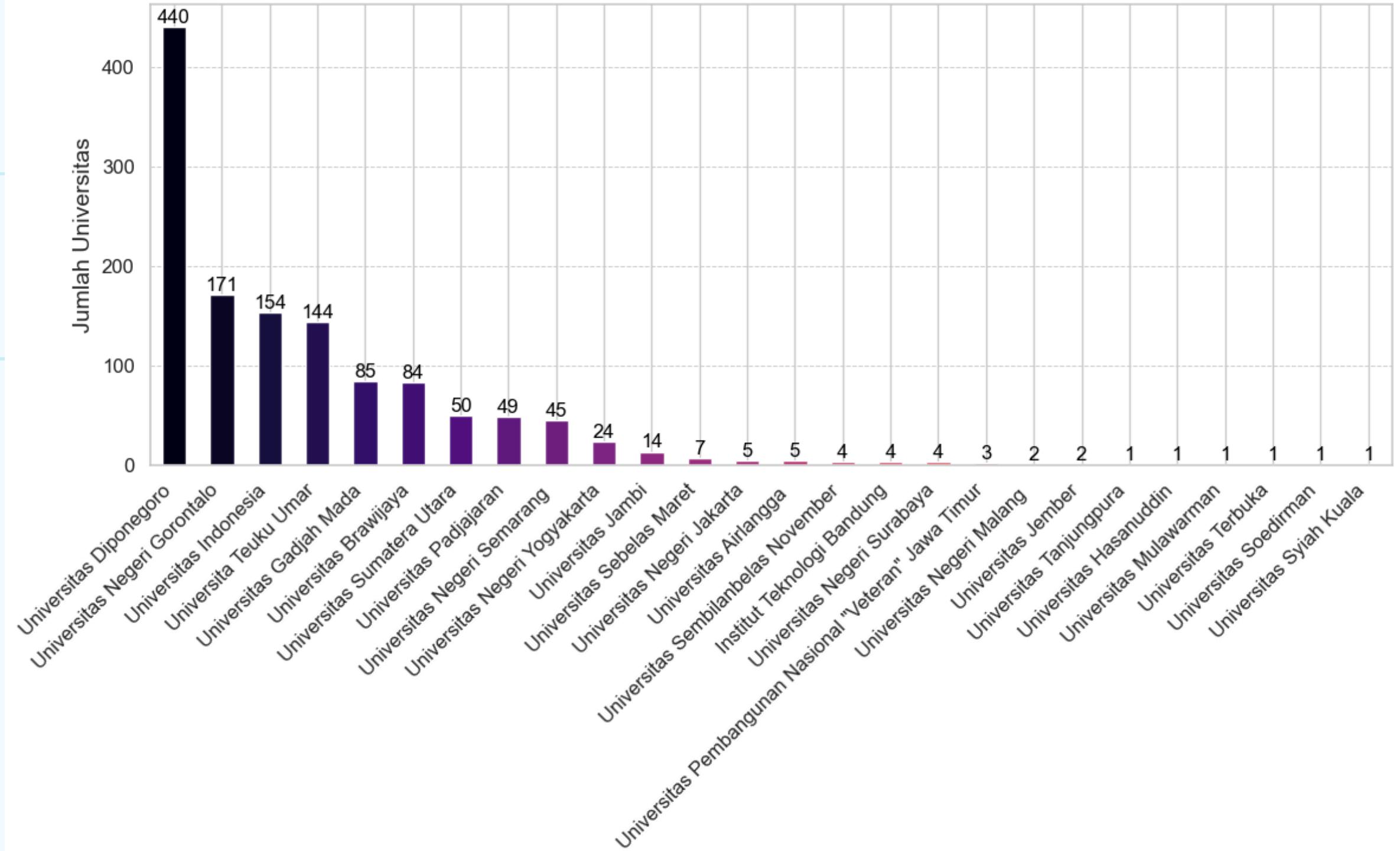
Universitas yang paling dibicarakan tentang KIPK

Jumlah Penyebutan Universitas pada Setiap Kategori



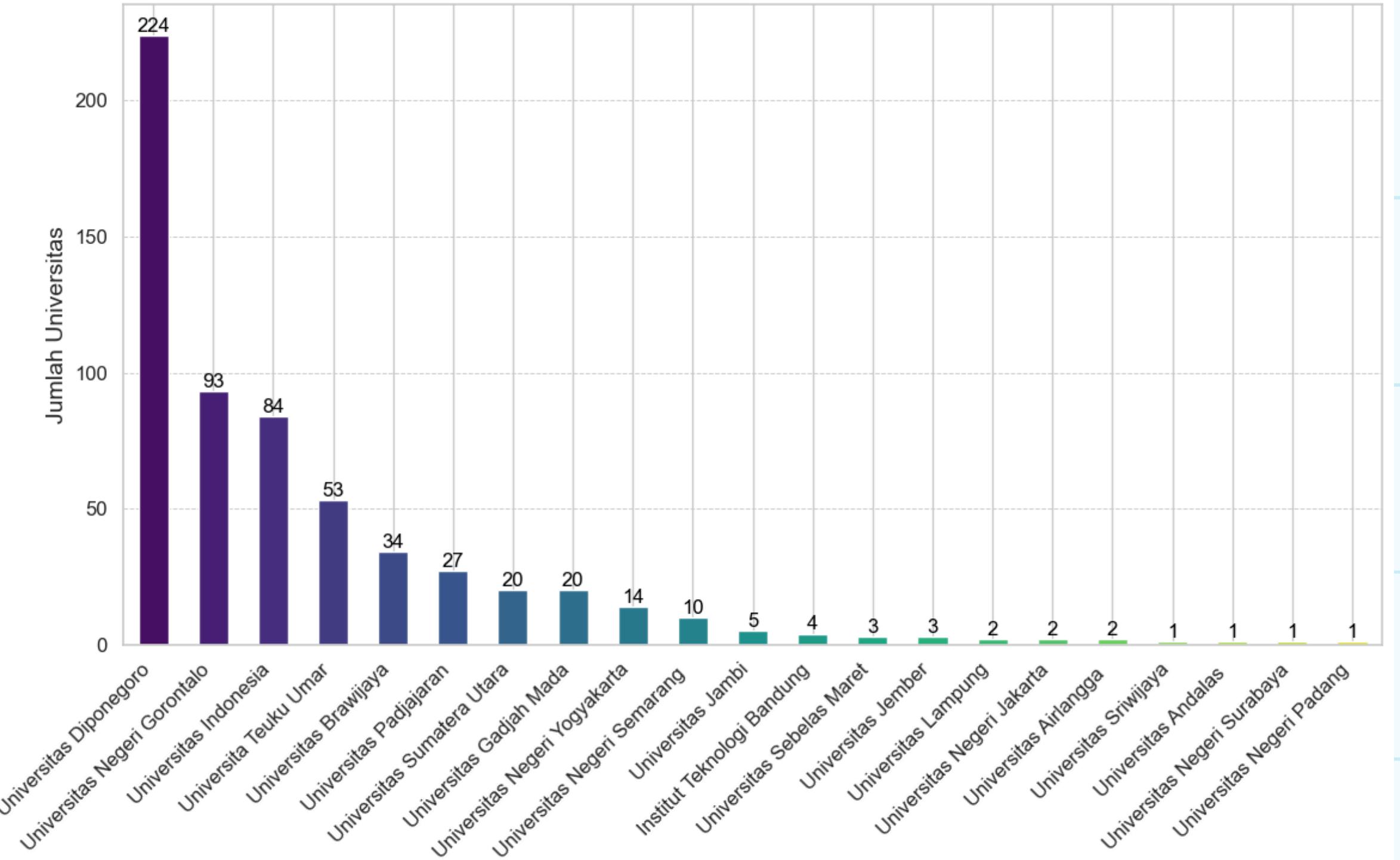
Universitas dengan Label Negative

Jumlah Universitas dengan Label Negatif pada Setiap Kategori

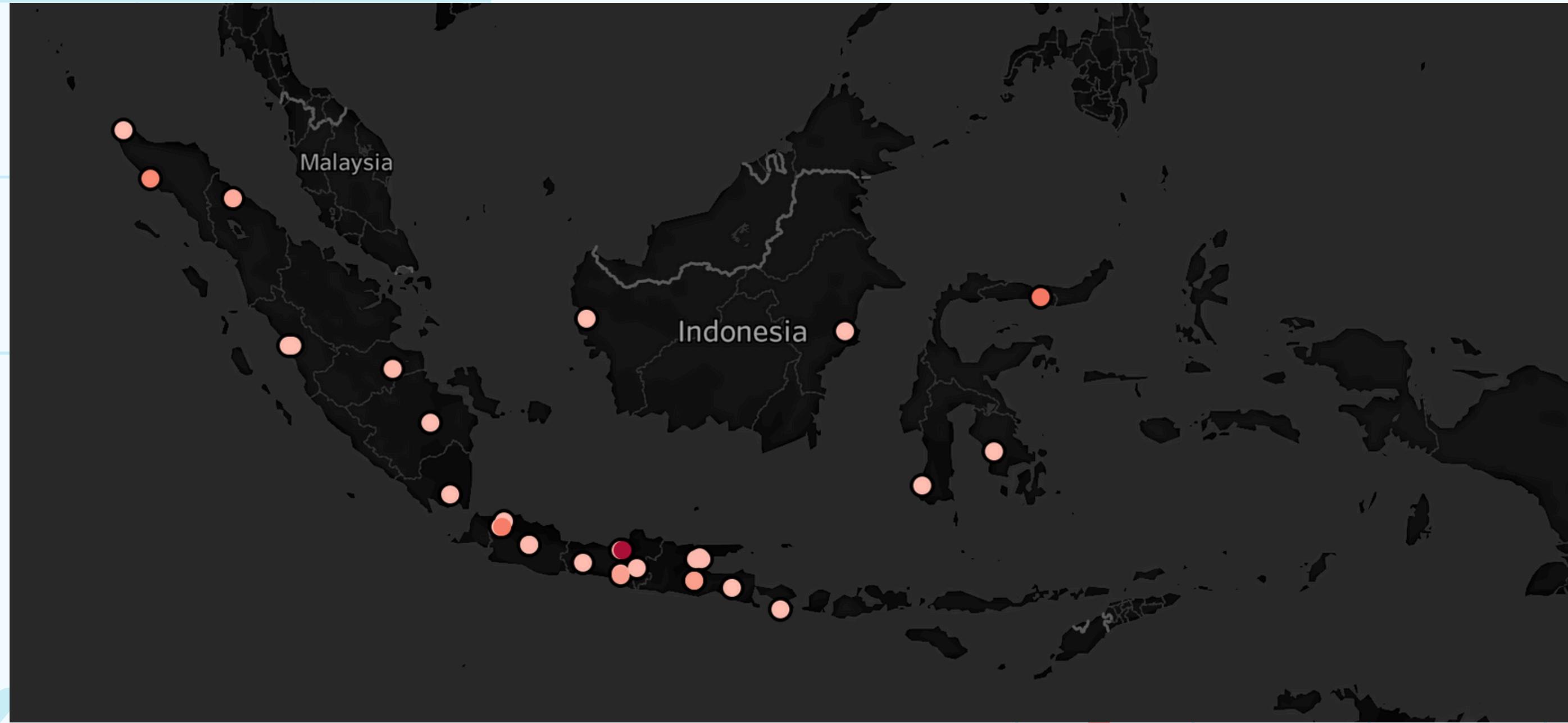


Universitas dengan Label Positive

Jumlah Universitas dengan Label Positif pada Setiap Kategori



Distribusi dengan geomaps



Hasil Analisis Sentimen

Berdasarkan hasil analisis sentimen, Universitas Diponegoro menjadi pusat perhatian di Twitter. Dalam rentang waktu dari 10 Juli 2023 hingga 30 April 2024, Universitas Diponegoro paling banyak mendapatkan mention dengan kata kunci 'KIPK'. Ini menunjukkan bahwa ada banyak diskusi dan minat yang signifikan terkait dengan Universitas Diponegoro dan program KIPK selama periode tersebut

Terima Kasih

Apakah ada pertanyaan?

