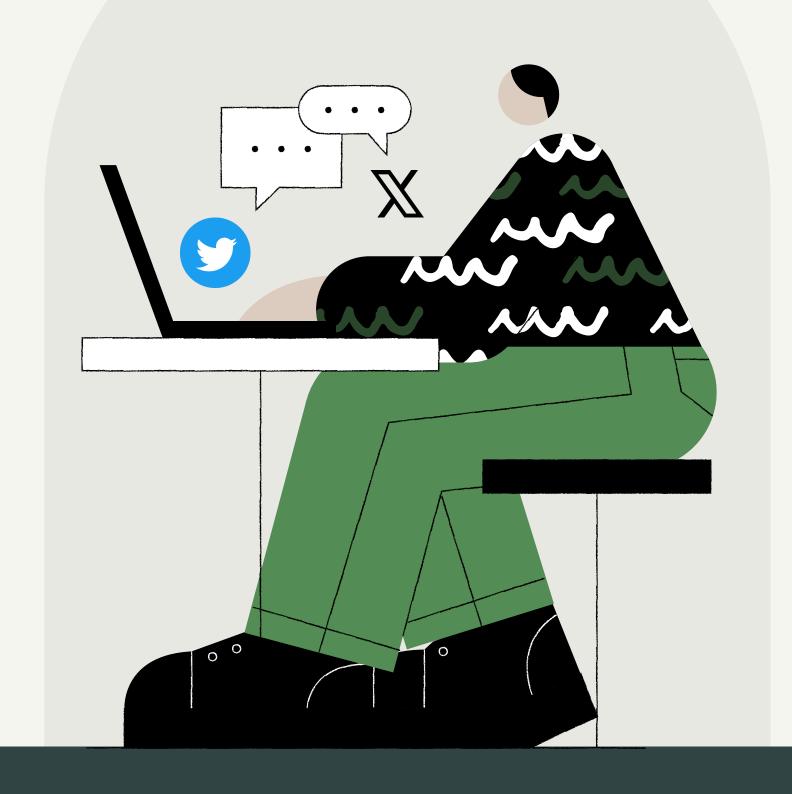


Analisis Total Karakter dan Kata pada Data Twitter menggunakan Descriptive Analytics

Data Science - Gold Challenge

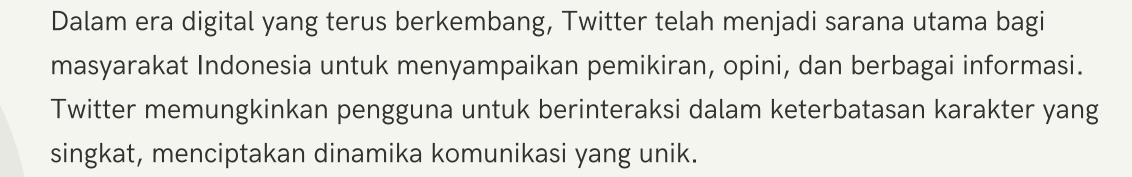


RAFI ARRIBAATH ALFARESY JANUARI 2024

Pendahuluan







Dengan keterbatasan 280 karakter per tweet, Twitter memaksa pengguna untuk merangkum pemikiran mereka secara singkat dan padat. Ini menciptakan bentuk ekspresi yang efektif dan langsung, memungkinkan penyebaran informasi dengan cepat dan tanpa kekakuan formal.

Dengan demikian, analisis total karakter atau kata dalam data Twitter di Indonesia tidak hanya membatasi diri pada aspek teknis, tetapi juga memberikan wawasan mendalam terhadap kreativitas, tren bahasa, respons terhadap peristiwa, dan dinamika sentimen dalam komunikasi singkat di dunia maya.

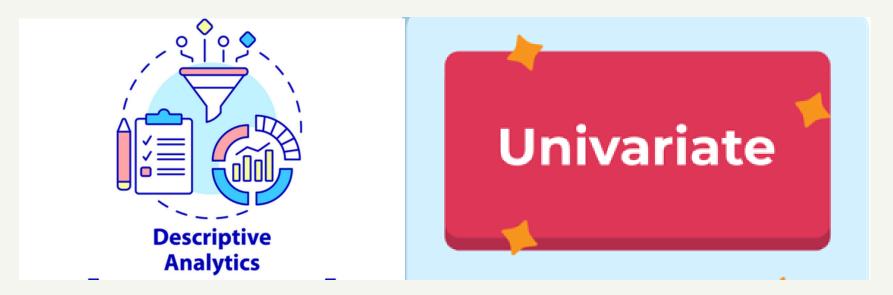
Metode Penelitian



```
@swag_from("docs/text_processing.yml", methods = ['POST'])
dapp.route('/text-processing', methods=['POST'])
  f text processing():
   text = request.form.get('text')
   text_clean = re.sub(r'[^a-zA-Z0-9]', ' ', text)
   json_response = {
        'status_code': 200,
        'description': "Teks yang sudah diproses",
        'data_raw':text,
        'data_clean': text_clean
   response_data = jsonify(json_response)
   return response data
@swag_from("docs/text_processing_file.yml", methods = ['POST']
@app.route('/text-processing-file', methods=['POST'])
ef text processing file():
   file = request.files.getlist('file')[0]
   df = pd.read_csv(file, encoding="ISO-8859-1")
   texts = df.Tweet.to_list()
   cleaned text = []
   for text in texts:
       cleaned_text.append(re.sub(r'[^a-zA-Z0-9]', '', text))
   json response = {
        'status_code' : 200,
        'description' : "Teks yang telah diproses",
        'data' : cleaned text,
   response_data = jsonify(json_response)
   return response data
    name == ' main ':
```

Proses Code Function Cleansing

Codingan dilakukan dalam Visual Studio
Code untuk membangun
API yang dapat melakukan data cleansing pada text dan juga pada text di dalam file atau bentuk CSV.



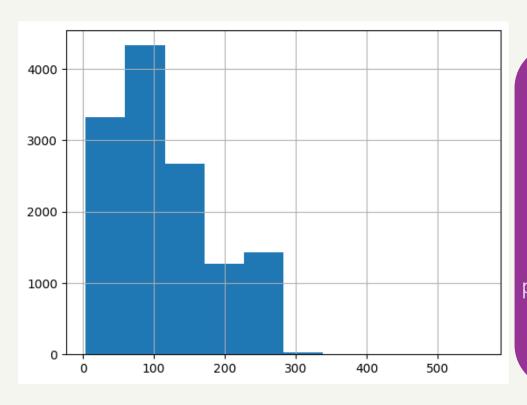
Pemilihan Metode Statistika dan EDA

Metode analisis yang dipakai dalam penelitian ini menggunakan descriptive analytics. Dikarenakan tujuannya adalah untuk menguraikan pola yang terdapat dalam data. Jenis analisis ini dianggap tepat karena difokuskan untuk mengidentifikasi kondisi data serta memahami pola-pola yang mungkin ada dalam data tersebut. Dan juga Univariate Analysis untuk menganalisis masingmasing dari data total karakter dan total kata.

Metode Penelitian

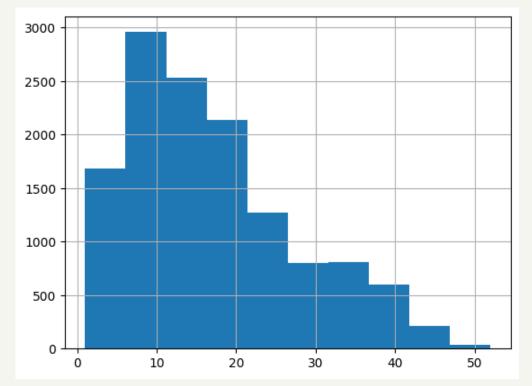
Visualisasi Data





Visualisasi 1

Visualisasi ini
menunjukkan bahwa
rata-rata panjang
karakter dari data tweet
pengguna adalah sebesar
60-110 karakter.



Visualisasi 2

Visualisasi ini menunjukkan bahwa rata-rata panjang kata dari data tweet pengguna adalah sebesar 6-15 kata.



Visualisasi 3

Visualisasi ini menampilkan kata-kata yang sering muncu di dalam semua tweet yang ada pada data, yakni "USER" "x9f", "xf0", "yang", "dan"



Visualisasi 4

Visualisasi ini menampilkan katakata yang sering muncul di dalam tweet yang termasuk kategori 'HS" dan "Abusive" yang ada pada data, yakni "berusaha", "melacak", "cowok", "disaat", "semua", "perhatian"

RAFI ARRIBAATH ALFARESY JANUARI 2024



Kesimpulan

Univariate Analysis

Visualisasi Data

Visualisasi Data 2

Dalam statistik deskriptif

Ditunjukkan bahwa data twitter memiliki outlier tapi tidak begitu signifikan. Di Dalam Visualisasi Data

Ditunjukkan bahwa total karakter dan total kata memiliki panjang 60-110 karakter dan 6-15 kata.

Lanjutan

Ditunjukkan bahwa kata yang sering muncul pada semua tweet adalah "USER", "x9f", "xf0", "yang", "dan", kemudian katakata yang sering muncul di dalam tweet yang termasuk kategori "HS" dan "Abusive" yang ada pada data, adalah "berusaha", "melacak", "cowok", "disaat", "semua", "perhatian"