

Analisis Pemilihan Kata Pengguna Twitter Dengan Metode Descriptive Analytic

Data Science - Gold Challenge

Jafei Wuysang

Pendahuluan

- Peran media sosial Twitter atau X di era digital terletak pada kemampuannya memfasilitasi komunikasi secara real-time, jangkauan global, potensi pemasaran, dan perannya dalam gerakan sosial dan politik.
- Karakteristik unik dari platform media sosial ini menjadikannya alat yang ampuh bagi individu, bisnis, organisasi, dan komunitas untuk terhubung, berbagi informasi, dan memberikan dampak pada skala global.
- Sebagai salah satu negara dengan jumlah pengguna internet yang tinggi, menurut data dari We Are Social, Indonesia memiliki lebih dari 27 juta pengguna Twitter.

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.

```
#TEXT PROCESSING -----
@swag_from("docs/text_processing.yml", methods = ['POST'])
@app.route('/text-processing', methods=['POST'])
def 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

#TEXT PROCESSING FILE-----
@swag_from("docs/text_processing_file.yml", methods = ['POST'])
@app.route('/text-processing-file', methods=['POST'])
def 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

if __name__ == '__main__':
    app.run()
```

Pemilihan Metode Statistika dan EDA

Penelitian ini menggunakan metode analisis descriptive analytics, yang bertujuan 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.

Selain itu, juga digunakan metode Univariate Analysis untuk menganalisis masing-masing dari data total karakter dan total kata.

Kesimpulan

- Dalam statistik deskriptif ditunjukkan bahwa data twitter memiliki outlier tapi tidak begitu signifikan.
- Visualisasi data menunjukkan bahwa total karakter dan total kata memiliki panjang 60-110 karakter dan 6-15 kata.
- Kata yang sering muncul pada semua tweet adalah "USER", "x9f", "xf0", "yang", "dan".
- Kata-kata yang sering muncul di dalam tweet yang termasuk kategori "HS" dan "Abusive" yang ada pada data, adalah "berusaha", "melacak", "cowok", "disaat", "semua", "perhatian"





THANK YOU