



PERBANDINGAN METODE MACHINE LEARNING UNTUK SENTIMEN ANALISIS REVIEW PENJUALAN PRODUK DI TOKOPEDIA

Muhammad Reza

2019470055

Teknik Informatika



Latar Belakang

Terdapat opini atau sentimen di ulasan pelanggan yang sudah membeli barang di toko online



penelitian ini melakukan analisa sentimen di e-commerce Tokopedia mengenai ulasan pelanggan di beberapa produk, yaitu elektronik, pakaian menggunakan metode machine learning supervised decision tree, k nearest neighbor, naives bayes



Rumusan Masalah

01.

Rumusan Satu

Bagaimana cara mengolah data teks untuk melakukan analisis sentimen, melakukan analisis sentimen dengan menghasilkan positif, nilai negatif, dan nilai netral, serta cara melakukan pelabelan sentimen berdasarkan rating bintang 1 sampai 5?

02.

Rumusan Dua

Bagaimana hasil dan akurasi perbandingan melakukan komparasi machine learning untuk analisis sentimen berdasarkan kondisi positif, negatif, netral?



Tujuan Penelitian

01.



02.



03.





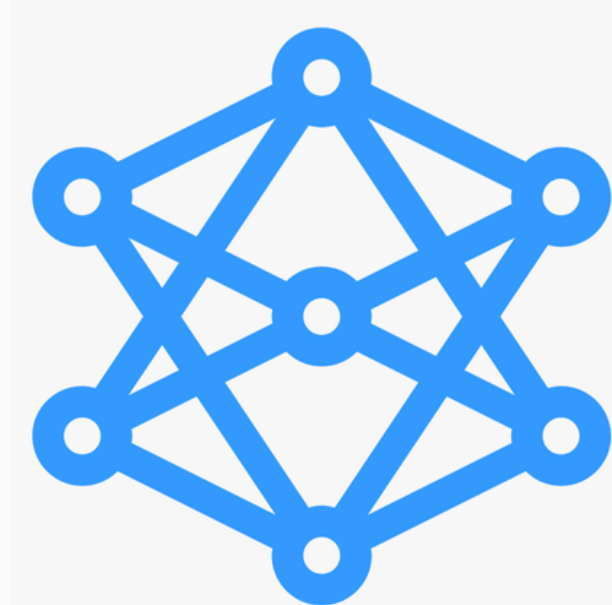
Metodologi Penelitian

01.



preprocessing data

02.



pemodelan

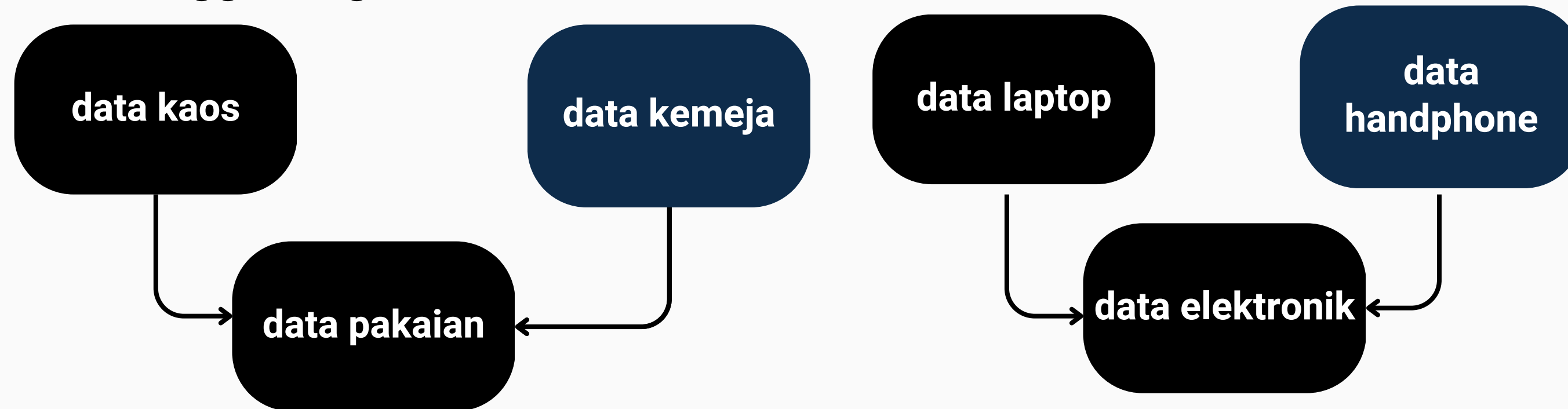
03.



evaluasi



01. Penggabungan



02. Pelabelan



03. Text preprocessing

Kiriman cepat
sampai. Bahan
kain agak
tebal. Ukuranny
a kurang lebar
dikit, berasa
bukan 52. Thanks

irim pesan sesuai
dgn pes kecewa
barang dtng tdk
sesuai pdhal
produk kenal



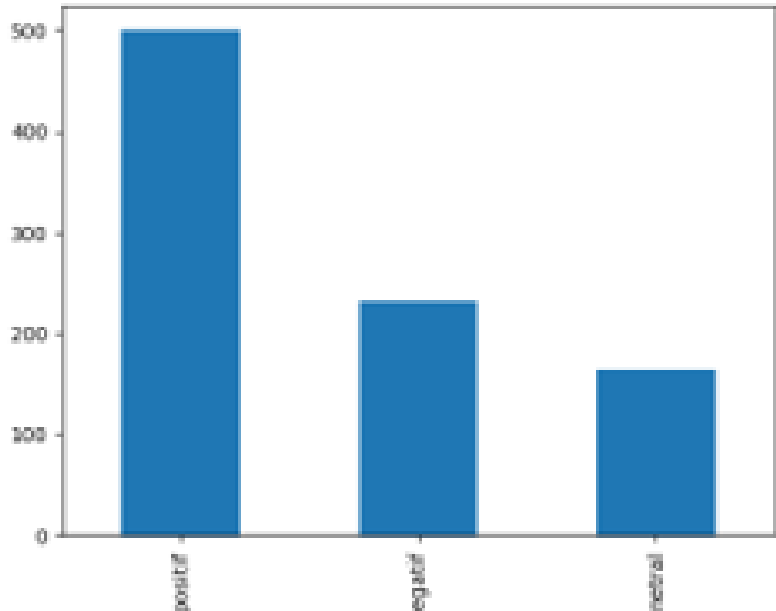
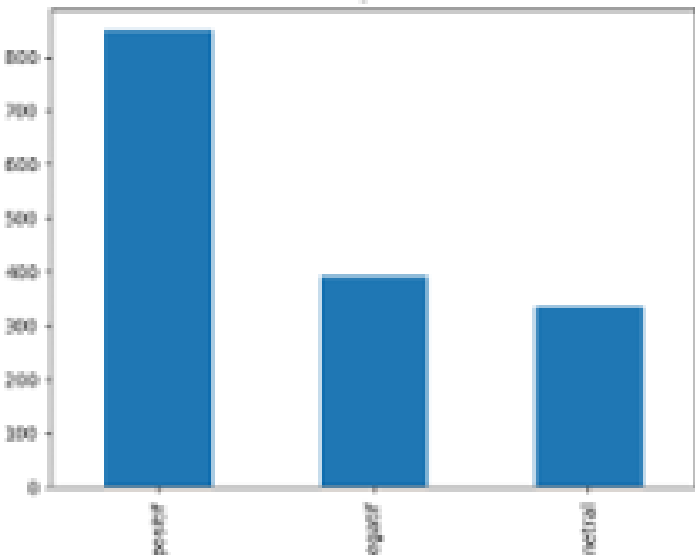
04. Pembobotan kata

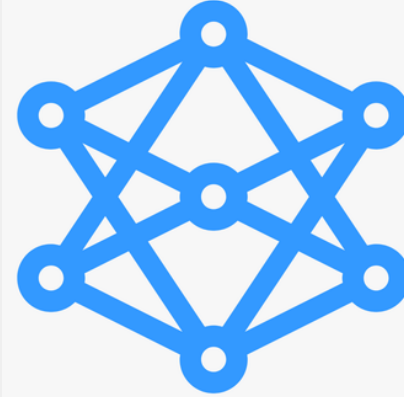
sesuai

$$IDF(sesuai) = \log_{10}(4/1) = 0,6020599913279624$$



02.1

Grafik data	Keterangan								
<p>Data Elektronik</p>  <table border="1"><thead><tr><th>Sentiment</th><th>Count</th></tr></thead><tbody><tr><td>positif</td><td>500</td></tr><tr><td>negatif</td><td>232</td></tr><tr><td>netral</td><td>163</td></tr></tbody></table>	Sentiment	Count	positif	500	negatif	232	netral	163	<p>Total data elektronik:</p> <ol style="list-style-type: none">1. Total label positif adalah 500 data.2. Total label negatif adalah 232 data.3. Total label netral adalah 163 data.
Sentiment	Count								
positif	500								
negatif	232								
netral	163								
<p>Data pakaian</p>  <table border="1"><thead><tr><th>Sentiment</th><th>Count</th></tr></thead><tbody><tr><td>positif</td><td>849</td></tr><tr><td>negatif</td><td>391</td></tr><tr><td>netral</td><td>335</td></tr></tbody></table>	Sentiment	Count	positif	849	negatif	391	netral	335	<p>Total data pakaian:</p> <ol style="list-style-type: none">4. Total label positif adalah 849 data.5. Total label negatif adalah 391 data.6. Total label netral adalah 335 data.
Sentiment	Count								
positif	849								
negatif	391								
netral	335								



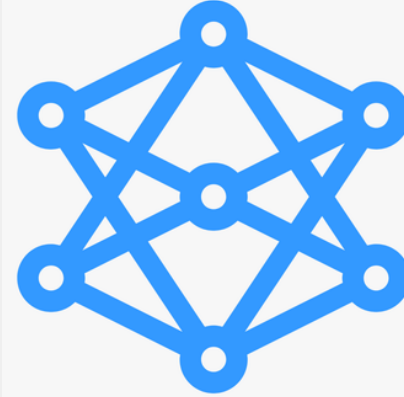
02

Naives bayes

k nearest neighbor

k=[1,5]

Decision tree



02.1

10% data uji

90% data latih

20% data uji

80% data latih

30% data uji

70% data latih

40% data uji

60% data latih



Evaluasi Data pakaian



03.1

	Naives bayes	Decision tree	k nearest neighbor k=5
evaluasi	30% data test dan 70% data latih	10% data test dan 90% data latih	10% data test dan 90% data latih
accuracy	37%	71%	65%
recall	37%	71%	65%
precision	53%	70%	63%



Evaluasi Data elektronik



03.2

	Naives bayes	Decision tree	k nearest neighbor k=4
evaluasi	40% data test dan 60% data latih	10% data test dan 90% data latih	10% data test dan 90% data latih
accuracy	41%	64%	67%
recall	41%	64%	67%
precision	53%	63%	65%

KESIMPULAN DAN SARAN



Kesimpulan

- Data yang diambil sebanyak 1575 data untuk data pakaian dan 895 data untuk data elektronik
- data pakaian menghasilkan label positif sebanyak 849 data, label netral sebanyak 391 data, label negatif sebanyak 335 data
- data elektronik menghasilkan label positif sebanyak 500 data, label netral sebanyak 168 data, label negatif sebanyak 232 data.

Hasil terbaik dari akurasi

- Menggunakan decision tree untuk data pakaian (10% data uji 90% data latih)
- Menggunakan k nearest neighbor untuk data elektronik (10% data uji 90% data latih nilai $K=4$)

Saran

- Penelitian ini berpengaruh terhadap banyaknya data uji dan data latih pada saat menghasilkan evaluasi berupa nilai akurasi, recall, precision.