PENGARUH MATRIKS FILTERS, FRAMEWORKS, DAN PRE TRAINING PADA PERBAIKAN PERFORMA TRAINING METODE DEEP LEARNING: CNN UNTUK ANALISIS SENTIMEN

Moch. Ari Nasichuddin
(ari.n@mail.ugm.ac.id)

Pembimbing I: Teguh Bharata Adji, S.T., M.T., M.Eng., Ph.D.
Pembimbing II: Widyawan, S.T., M.Eng., Ph.D.

Program Studi S2 Teknologi Informasi
Departemen Teknik Elektro dan Teknologi Informasi, Fakultas Teknik, Universitas Gadjah Mada

1 INTISARI

Pendekatan memakai metode Deep Learning memberikan hasil yang bagus dalam implementasi di berbagai bidang, tak terkecuali di bidang Analisis Sentimen. Salah satu metode Deep Learning yakni CNN, mampu memberikan akurasi yang bagus di beberapa penelitian sebelumnya. Namun masih ada beberapa bagian dalam proses training yang dapat diperbaiki untuk meningkatkan tingkat akurasi dan waktu pelatihannya. Penelitian ini mencoba melakukan perbaikan pada model CNN untuk mendapatkan akurasi dan waktu proses yang terbaik. Untuk memenuhi tujuan tersebut maka diterapkan beberapa tahap yakni dengan melakukan tuning pada penggunaan filter size, penggunaan library, dan penggunaan pre-training. Hasil simulasi menunjukkan penggunaan filter size berukuran lebih kecil dan pre-training word2vec memberikan akurasi lebih besar mencapai daripada beberapa penelitian sebelumnya.

2 PENDAHULUAN

Ketersediaan data berbentuk teks di era Big Data sangat besar

→ Metode yang sudah ada (Naïve Bayes, Maximum Entropy, dan Support Vector Machine (SVM)) memiliki **kelemahan** ketika menangani **data yang jumlahnya besar.**

→ Metode Deep Learning (CNN) **mampu menangani data yang jumlahnya besar**, namun memerlukan **waktu training lama** dan **arsitektur yang sesuai** agar menghasilkan akurasi yang terbaik

→ Dilakukan tuning pada metode Deep Learning (CNN) agar mendapatkan waktu training dan akurasi yang terbaik.

3 TUJUAN

Penelitian ini memiliki tujuan utama meningkatkan performa metode Deep Learning: CNN yang diimplementasikan pada kasus analisis sentimen. Adapun secara terperinci bisa dilihat sebagai berikut:

1. Melakukan peningkatan akurasi training dengan melakukan tuning pada penggunaan filter size, pre-training, framework yang dipakai.
2. Melakukan peningkatan time processing ketika melakukan training

4 TINJAUAN PUSTAKA

Penulis	Tahun	Metode	Data	Perbandingan
Kim Y.	2014	CNN dengan pre-training word2vec	Movie Review dan Twitter	Ukuran Matriks Filters pada metode CNN berbeda. Proses training memakai frameworks Tensorflow dan Theano.
Hassan A. et al.	2017	CNN dan LSTM	Movie Review	Metode yang dipakai hanya CNN. Proses pre-training menggunakan one hot encoding dan word2vec.
Pasi G. K., et al.	2017	CNN (1 convolution layer)	Movie Review	Ukuran Matriks Filters pada metode CNN berbeda. Proses training memakai frameworks Tensorflow dan Theano.
Ouyang X. et al.	2015	CNN (3 convolution layer)	Movie Review	Metode CNN yang dipakai hanya 1 convolution layer. Proses pre-training menggunakan one hot encoding dan word2vec.

5 METODOLOGI

Studi Literatur
↓
Menganalisis Aspek untuk Tuning CNN
↓
Menyiapkan Dataset
↓
Menyiapkan Parameter Setting, Pre-training, dan Framework
↓
Percobaan Training
↓
Analisis Akurasi dan Waktu Training
↓
Kesimpulan dan Saran

6 HASIL

Parameters Setting	Framework	Akurasi	Waktu Pemrosesan
I	TF	74,1%	**2m4s**
	TF + W2V	79,4%	14m24s
	TH + W2V	**80,9%**	53m
II	TF	74,8%	**8m43s**
	TF + W2V	73,7%	9m22s
	TH + W2V	**80,45%**	6m66s
III	TF	78,1%	9m45s
	TF + W2V	78,9%	**8m34s**
	TH + W2V	**80,2%**	6m66s
IV	TF	79,3%	**66m48s**
	TF + W2V	**81,10%**	72m57s
	TH + W2V	80,67%	110m9s

Ket: **TF** = Tensorflow **TH** = Theano **W2V** = Word2vec

7 KESIMPULAN

Dari hasil percobaan yang diusulkan dapat disimpulkan beberapa hal. Pertama, arsitekur CNN dengan Convolution Layer terlalu banyak dapat menurunkan akurasi jika menggunakan dataset dengan baris data yang kecil. Oleh karena itu untuk menentukan arsitektur CNN perlu mempertimbangkan banyaknya data yang ada. Kedua, ukuran Matriks Filter yang lebih kecil memberikan tingkat akurasi yang lebih tinggi daripada ukuran yang lebih besar, hal ini karena kata yang diolah akan semakin banyak. Ketiga, penggunaan pre-training word2vec memberikan peningkatan akurasi yang lebih baik karena mempertimbangkan posisi dan konteks dari sebuah kata di sebuah kalimat. Hal ini berbeda dengan one hot encoding yang hanya berdasarkan posisi kata di sebuah kalimat saja. Untuk penelitian ke depan, pengembangan penelitian dengan memanfaatkan arsitektur CNN yang lebih kompleks sangat diperlukan untuk merangani data yang besar.