

Nama :Satria Nur Saputro

Kelas : Kangae

Institute : Institute Teknologi Telkom Purwokerto

Domain : NLP

TUGAS ANALISIS

Link collab : https://colab.research.google.com/drive/1_15WcYKX9o_3OcvmE1ZqE6i-xFyy4w5s?usp=sharing

Buat resume perbandingan metode tradisional ML, LSTM, dan fine-tuning BERT pada data klasifikasi SMS Spam.

- Gunakan metrics akurasi untuk melihat performa metode diatas;

Metode	Akurasi	Waktu(s)	Epochs
Naïve Bayes	92,5%	0,022	-
SGD	91,5%	0,083	-
LSTM	86%	81	9
LSTM + Word2vec	82,53%	374	23
SimpleRNN + Word2vec	82,09%	335	8
BERT	94,65%	445	3

Catatan :

- Semua uji menggunakan proses split train 80% dan test 20%, random state 42
- Uji SGD menggunakan fine tuning : penalty = 12 dan max_iter = 5
- Penggunaan early stopping untuk mendapatkan hasil optimal dengan delta 4 monitor val_loss = min
- Uji BERT menggunakan fine_tuning learning rate = 5e-5
- Seluruh deep learning menggunakan batch_size = 2
- Seluruh penelitian menggunakan Google Collab run time GPU

- Analisa hasil performa setiap metode;

Berdasarkan metric akurasi yang dilakukan pada data uji/data testing baik metode machine learning dan deep learning dengan algoritma Multinomial NB, LSTM, dan BERT, metode BERT memiliki performa yang sangat baik dengan mendapatkan akurasi di atas 94% tanpa overfitting ataupun underfitting. Tetapi jika melihat waktu komputasi *training* Naïve Bayes memiliki waktu singkat 0,022 s

- metode mana yang lebih baik, dari segi komputasi dan performa? Jelaskan mengapa?

Untuk kasus dataset spam sms , ini menurut saya lebih baik menggunakan machine learning naïve bayes karena memiliki accurary dan juga waktu komputasi yang lebih singkat kurang dari satu detik, kemudian tidak banyak menggunakan resource besar dan mudah dideploy, tetapi jika mengejar accuracy tinggi *deep learning* BERT merupakan metode optimal untuk mendapatkan hasil terbaik dengan beberapa kekurangan yaitu jumlah *resources* CPU dan *Memory* sangat besar yang perlu diperhatikan dan waktu komputasi yang lama. Maka dapat

disimpulkan metode machine learning dengan algoritma Multinomial NB bekerja lebih baik pada kasus klasifikasi SMS Spam.