



Biodata Perkuliahan Angket CSI Keluar

User: C14170054 - JEHEZKIEL HARDWIN TANDIJAYA

Hak Akses: Mahasiswa - INFORMATIKA

Login Terakhir: 01-07-2021 15:08:44

(Indonesia)

Revisi Skripsi Mahasiswa

Cetak BA Revisi

Nama : JEHEZKIEL HARDWIN TANDIJAYA
NRP : C14170054

Fakultas : TEKNOLOGI INDUSTRI
Jurusan : INFORMATIKA

Judul TA : Implementasi Web Scraping dan Klasifikasi Dalam Pembuatan Portal Berita Online

Dosen Penguji	
NIP	14001
Nama Pegawai	Henry Novianus Palit, S.Kom., M.Kom., Ph.D.
Email	hnpalit@peter.petra.ac.id

DAFTAR REVISI	
No	Revisi
1	Judul: Tidak perlu menyebutkan implementasi, karena impresinya skripsi hanya sebatas itu. Web scraping juga, menurut saya, tidak perlu disebutkan. Seharusnya BERT (sebagai teknologi utama) yang harus muncul dalam judul.
2	Latar belakang dan rumusan masalah: Karena dalam rumusan masalah sangat ditekankan klasifikasi teks bahasa Indonesia, seharusnya alasannya disinggung juga dalam latar belakang masalah.
3	Ruang lingkup: Poin terakhir, "Metode yang digunakan adalah BERT", tolong diperjelas metode untuk apa?
4	Subbab 2.1.2: Untuk persamaan (2.1) s/d (2.4) berikan penjelasan masing-masing variabel yang digunakan, misal TP, FP, TN, FN dst. Dituliskan secara konsisten, precision atau presisi?
5	Subbab 2.2: Paragraf-paragrafnya dibuat rata dengan batas kiri halaman?
6	Subbab 5.1: Pengujian diarahkan untuk menjawab perumusan masalah (mengukur seberapa jauh solusi yang diusulkan dapat menjawab perumusan masalah). Apakah pengujian web scraping menjawab perumusan masalah?
7	Subbab 5.2: Dijelaskan dulu model BERT apa saja yang digunakan, mengapa model-model tersebut yang dipilih, dst. Mengapa learning rate nya diset berbeda antara bert-base dan bert-base-multilingual-uncased? Sebaiknya disamakan saja, dari 1e-5 s/d 5e-5. Mengapa epoch dibuat konstan (diset 3), tapi pada bagian selanjutnya ada pengujian epoch (2 s/d 4)? Lebih baik epoch-nya menjadi salah satu variabel yang diuji. Koreksi batch size nya menjadi 16.
8	Subbab 5.2: Sebelum pengujian, jelaskan dulu jumlah keseluruhan data, lalu bagaimana membagi data untuk training, testing, dan validasi. Jelaskan juga jumlah data per class/kategori, sebelum dan sesudah dibagi ke dalam training, testing, dan validasi. Setelah pengujian, untuk setiap konfigurasi yang dipilih, berikan evaluasi dan alasan/justifikasi.
9	Subbab 5.3: Gunakan confusion matrix yang sesuai untuk klasifikasi multi-class. Berikan analisa terhadap hasilnya.
10	Bab 7: Beberapa referensi perlu didetailkan. Misalnya: Allahyari et al. (2017), Ouatik et al. (2021), Paul dan Saha (2020), dst.
11	Program: Berikan multi-tag pada setiap berita. Tentukan threshold untuk memilih tag-tag yang relevan.