Manual Book

“Indobertplag”

Aplikasi Cek Plagiarisme Karya Ilmiah Berbasis Web

Indobertplag adalah sebuah aplikasi cek plagiarisme berbasis web yang khusus digunakan untuk karya ilmiah. Aplikasi ini mengeluarkan tingkat kemiripan antara teks input dengan teks-teks yang terdaftar pada database. Dari nilai kemiripan tersebut, sebuah teks dapat diklasifikasi sebagai:

* Non plagiat (0 - 9.9%),
* plagiat ringan (10 - 29.9%),
* Plagiat berat (30 - 100%).

Aplikasi ini menggunakan model IndoBERT, sebuah model LLM berbasis arsiektur transformers, dan metode *cosine similarity* untuk menghitung tingkat kemiripan.

# **Instalasi**

Aplikasi ini dijalankan pada komputer lokal, dan dirancang dengan menggunakan framework Django (berbasis Python) dan database PostgreSQL. Maka sebelum dapat menjalankan aplikasi, perlu melakukan instalasi dan setting.

## **Download**

Sebelum dapat memulai instalasi, download dan instalasi yang perlu dilakukan adalah:

* [Python versi 3.10.8](https://www.python.org/downloads/release/python-3108/)
* [PostgreSQL versi 17.0](https://www.postgresql.org/download/)
* [pgAdmin 4](https://www.pgadmin.org/download/)
* IDE seperti [Visual Studio Code versi 19.5](https://code.visualstudio.com/docs/?dv=win64user)

## **Instalasi framework Django dan library**

Program indobertplag terdiri dari:

* Folder indobertplag,
* File requirements.txt, dan
* File README.md

File README.md adalah manual ringkas. Lakukan instalasi Django dan library-library yang terdapat dalam requirements.txt:

* Buka folder program dalam IDE seperti Visual Studio Code
* Buka *terminal* pada IDE (untuk Visual Studio Code dapat menggunakan *shortcut* CTRL + Shift + `),
* Buat sebuah virtual environment dengan menjalankan 2 *command* berikut:

python -m venv Venv

.\Venv\Scripts\Activate

A screenshot of a computer

Description automatically generated

* Untuk menginstall semua *library* yang digunakan program, jalankan *command* berikut:

pip install -r requirements.txt

A screenshot of a computer

Description automatically generated

*Command* tersebut akan menginstall *library-library* yang tercatat dalam requirements.txt secara otomatis, yang terdiri dari:

* [asgiref versi 3.8.1](https://pypi.org/project/asgiref/3.8.1/)
* [certifi versi 2024.8.30](https://pypi.org/project/certifi/2024.8.30/)
* [charset-normalizer versi 3.4.0](https://pypi.org/project/charset-normalizer/3.4.0/)
* [colorama versi 0.4.6](https://pypi.org/project/colorama/0.4.6/)
* [Django versi 5.1.2](https://pypi.org/project/Django/5.1.2/)
* [filelock versi 3.16.1](https://pypi.org/project/filelock/3.16.1/)
* [fsspec versi 2024.9.0](https://pypi.org/project/fsspec/2024.9.0/)
* [huggingface-hub versi 0.25.2](https://pypi.org/project/huggingface-hub/0.25.2/)
* [idna versi 3.10](https://pypi.org/project/idna/3.10/)
* [Jinja2 versi 3.1.4](https://pypi.org/project/Jinja2/3.1.4/)
* [joblib versi 1.4.2](https://pypi.org/project/joblib/1.4.2/)
* [MarkupSafe versi 3.0.1](https://pypi.org/project/MarkupSafe/3.0.1/)
* [mpmath versi 1.3.0](https://pypi.org/project/mpmath/1.3.0/)
* [networkx versi 3.4.1](https://pypi.org/project/networkx/3.4.1/)
* [numpy versi 2.1.2](https://pypi.org/project/numpy/2.1.2/)
* [packaging versi 24.1](https://pypi.org/project/packaging/24.1/)
* [psycopg2-binary versi 2.9.9](https://pypi.org/project/psycopg2-binary/2.9.9/)
* [PyMuPDF versi 1.24.11](https://pypi.org/project/PyMuPDF/1.24.11/)
* [PyYAML versi 6.0.2](https://pypi.org/project/PyYAML/6.0.2/)
* [regex versi 2024.9.11](https://pypi.org/project/regex/2024.9.11/)
* [requests versi 2.32.3](https://pypi.org/project/requests/2.32.3/)
* [safetensors versi 0.4.5](https://pypi.org/project/safetensors/0.4.5/)
* [scikit-learn versi 1.5.2](https://pypi.org/project/scikit-learn/1.5.2/)
* [scipy versi 1.14.1](https://pypi.org/project/scipy/1.14.1/)
* [sqlparse versi 0.5.1](https://pypi.org/project/sqlparse/0.5.1/)
* [sympy versi 1.13.3](https://pypi.org/project/sympy/1.13.3/)
* [threadpoolctl versi 3.5.0](https://pypi.org/project/threadpoolctl/3.5.0/)
* [tokenizers versi 0.20.1](https://pypi.org/project/tokenizers/0.20.1/)
* [torch versi 2.4.1](https://pypi.org/project/torch/2.4.1/)
* [tqdm versi 4.66.5](https://pypi.org/project/tqdm/4.66.5/)
* [transformers versi 4.45.2](https://pypi.org/project/transformers/4.45.2/)
* [typing\_extensions versi 4.12.2](https://pypi.org/project/typing-extensions/4.12.2/)
* [tzdata versi 2024.2](https://pypi.org/project/tzdata/2024.2/)
* [urllib3 versi 2.2.3](https://pypi.org/project/urllib3/2.2.3/)

## **Model dan *cosine similarity***

* Library sudah terinstall pada penjalanan fungsi (pip install -r requirements.txt)
* Model tidak di download dalam bentuk file secara manual, tetapi langsung menggunakan library huggingface. Model dinyatakan dalam file utils.py (indobertplag -> plagiarism -> utils.py).
* Running pertama kali [akan mendownload model ke cache lokal](https://huggingface.co/docs/transformers/en/installation) yang disimpan dalam C:\Users\username\.cache\huggingface\hub.
* Untuk menjalankan secara offline, daftarkan environment variable HF\_HUB\_OFFLINE=1.

Aplikasi ini menggunakan model IndoBERT yang telah di *fine-tune* oleh LazarusNLP yaitu [all-indobert-base-v4](https://huggingface.co/LazarusNLP/all-indobert-base-v4), yang di-install secara otomatis oleh library [huggingface-hub versi 0.25.2](https://pypi.org/project/huggingface-hub/0.25.2/) dan digunakan library [transformers versi 4.45.2](https://pypi.org/project/transformers/4.45.2/).

A screen shot of a computer program

Description automatically generated

Metode evaluasi *cosine similarity* adalah sebuah fungsi yang dapat dipanggil dan terdapat dalam [scikit-learn versi 1.5.2](https://pypi.org/project/scikit-learn/1.5.2/) yang juga sudah dinyatakan dalam requirements.txt.

## **Setting database**

Indobertplag menggunakan PostgreSQL untuk menyimpan data teks referensi dan data login untuk administrasi. Pertama, buka pgAdmin 4 dan buat sebuah database:

* Pada bagian kiri layer pgAdmin, klik kanan pilihan “Databases”, lalu hover pilihan “Create”
* Pilih/klik kiri pilihan “Database”:

A screenshot of a computer

Description automatically generated

* Akan muncul sebuah pop-up/modal/halaman kecil. Berikan nama indobertplag\_db, lalu tekan Save.

A screenshot of a computer

Description automatically generated

* Setelah itu, daftarkan database tersebut kedalam settings.py (indobertplag -> plagcheck -> settings.py). Isi pilihan “DATABASE” sehingga terlihat seperti berikut:

A screenshot of a computer

Description automatically generated

* Untuk ‘USER’ dan ‘PASSWORD’ dapat diisi oleh USER dan PASSWORD untuk PostgreSQL yang dibuat data pertama kali membuka pgAdmin 4. Pilihan lain sebaiknya sama persis.
* Migrasi perubahan-perubahan yang dilakukan dengan menggunakan *command* di terminal IDE. Pertama, pindah ke folder indobertplag:

cd indobertplag

* Setelah itu, jalankan fungsi migrasi Django:

python manage.py migrate

## **Buat super user**

Proyek Django dan PostgreSQL akan terhubung. Untuk menambahkan user admin, lakukan langkah-langkah berikut:

* Pada terminal IDE, jalankan fungsi:

python manage.py createsuperuser

* Terminal akan meminta untuk mengisi username, email, dan password. Username dan password akan digunakan untuk halaman login.

# **Menjalankan Aplikasi**

Untuk menjalankan aplikasi, pastikan bahwa terminal berada dalam folder indobertplag (C:\Users\user\folder\_program\indobertplag).

* Apabila masih di dalam folder program gunakan *command* berikut:

cd indobertplag

* Jalankan aplikasi dengan menggunakan *command*:

python manage.py runserver

A screenshot of a computer

Description automatically generated

* Setelah runserver, buka link tersebut di sebuah web browser:

Starting development server at <http://127.0.0.1:8000/>

A screenshot of a computer

Description automatically generated

# **Aplikasi**

## **Upload plagiarisme**

Halaman pertama yang akan buka adalah tampilan utama:

A screenshot of a computer

Description automatically generated

Halaman ini adalah tempat upload untuk mengecek kemiripan sebuah dokumen yang suspek.

Untuk menggunakannya, tekan tombol “browse”, pilih file PDF yang ingin dicek, lalu tekan tombol biru besar dengan tulisan “Upload PDF Anda”.

Setelah aplikasi selesai memproses teks input, maka pengguna akan dialihkan ke tampilan hasil cek plagiarisme:

A screenshot of a computer

Description automatically generated

Antarmuka ini menunjukkan hasil cek plagiarisme dalam nilai persentase dan hasil klasifikasi. Halaman ini terdapat 3 pilihan untuk pengguna, yaitu:

* Membuka file dokumen teks yang di-input: Tekan nama file (teks biru dibawah “Hasil Cek Plagiarisme”).
* Kembali ke antarmuka utama: Tekan teks “Upload PDF Lain” di paling bawah layar.
* Mengecek teks referensi yang ada dalam database: Tekan tombol “Referensi” pada kiri atas web.

Dalam penggunaan pertama, apabila langsung mengecek sebuah karya ilmiah, halaman tersebut akan kosong, karena belum ada teks referensi dalam database. Untuk itu, teks referensi ditambahkan dengan mengakses halaman tambah referensi melalui langkah-langkah berikut.

## **Referensi**

Menekan tombol “REFERENSI” akan memindah pengguna ke halaman login:

A screenshot of a computer

Description automatically generated

Login dengan menggunakan data yang dipilih dalam proses [membuat super user](#_Buat_super_user). Apabila sukses, maka akan muncul daftar teks referensi yang ada.

A screenshot of a computer

Description automatically generated

Halaman ini juga digunakan untuk menghapus teks referensi yang ada. Tentunya, dalam penggunaan pertama, tabel tersebut akan kosong, karena database masih kosong. Untuk mengisi database referensi, tekan “Tambah Teks Referensi”.

A screenshot of a computer

Description automatically generated

Halaman ini digunakan untuk menambah teks referensi kedalam database. Untuk penggunaan:

* Kotak judul dan penulis diisi dengan informasi karya ilmiah.
* Upload file PDF yang sesuai dengan tombol “Browse”.
* Tekan tombol “Simpan Referensi” untuk menyimpan data.

Setelah selesai diproses, maka pengguna akan dialihkan ke antarmuka daftar referensi. Pengguna dapat Kembali ke antarmuka utama dengan menekan tombol “Cek Plagiarisme”.