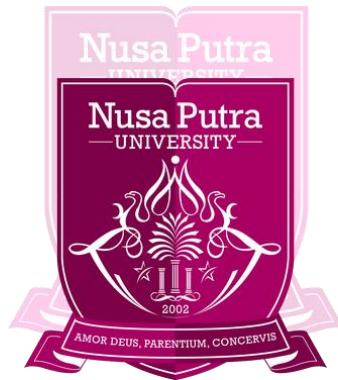


**IMPLEMENTASI SIAMESE NEURAL NETWORK UNTUK DETEKSI
KEMIRIPAN KODE : STUDI KASUS PENILAIAN TUGAS KODING
MAHASISWA NUSA PUTRA**

SKRIPSI

Nama : Deni Setiawan

Nama : Ilham Nurabduljabbar K NIM : 20210040011



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK, KOMPUTER DAN DESAIN

UNIVERSITAS NUSA PUTRA

SUKABUMI

AGUSTUS 2025

**IMPLEMENTASI SIAMESE NEURAL NETWORK UNTUK DETEKSI
KEMIRIPAN KODE : STUDI KASUS PENILAIAN TUGAS CODING
MAHASISWA NUSA PUTRA**

SKRIPSI

*Diajukan Untuk Memenuhi Salah Satu Syarat Dalam Menempuh Gelar
Sarjana Komputer Pada Program Studi Teknik Informatika*

Nama : Deni Setiawan

NIM : 20210040172

Nama : Ilham Nurabduljabbar K

NIM : 20210040011



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK, KOMPUTER DAN DESAIN
UNIVERSITAS NUSA PUTRA
SUKABUMI
AGUSTUS 2025**

PERNYATAAN PENULIS

JUDUL : IMPLEMENTASI SIAMESE NEURAL NETWORK UNTUK
DETEKSI KEMIRIPAN KODE: STUDI KASUS PENILAIAN
TUGAS CODING MAHASISWA NUSA PUTRA

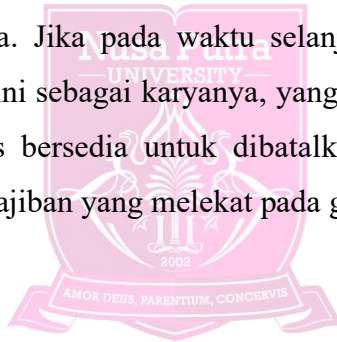
NAMA : DENI SETIAWAN

NIM : 20210040172

NAMA : ILHAM NURABDULJABBAR KOMARA

NIM : 20210040011

”Saya menyatakan dan bertanggungjawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya kami sendiri kecuali cuplikan dan ringkasan yang masing-masing telah dijelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai dengan bukti- bukti yang cukup, maka penulis bersedia untuk dibatalkan gelar Sarjana Komputer beserta segala hak dan kewajiban yang melekat pada gelar tersebut”.



Sukabumi,.....

Materai

Deni Setiawan

Penulis 1

Materai

Ilham Nurabduljabbar Komara

Penulis 2

PENGESAHAN SKRIPSI

JUDUL : IMPLEMENTASI SIAMESE NEURAL NETWORK UNTUK
DETEKSI KEMIRIPAN KODE: STUDI KASUS PENILAIAN
TUGAS CODING MAHASISWA NUSA PUTRA

NAMA PENULIS 1 : DENI SETIAWAN

NIM PENULIS 1 : 20210040172

NAMA PENULIS 2 : ILHAM NURABDULJABBAR KOMARA

NIM PENULIS 2 : 20210040011

Skripsi ini telah diujikan dan dipertahankan di depan Dewan Penguji pada Sidang
Skripsi tanggal _____ Menurut pandangan kami, Skripsi ini memadai dari
segi kualitas untuk tujuan penganugerahan gelar Sarjana Komputer (S.Kom)

Sukabumi,.....Agustus 2025

Pembimbing I

Pembimbing II



Nugraha, M.Kom
NIDN. 0413098904

Ivana Lucia Kharisma, M.Kom
NIDN. 0429038002

Ketua Penguji

Ketua Program Studi Teknik Informatika

Anggun Fergina, M.Kom
NIDN. 0407029301

Ir. Somantri, S.T, M.Kom
NIDN. 0419128801

Plh Dekan Fakultas Teknik Komputer dan Desain

Ir. Paikun S.T.,M.T.,IPM., ASEAN Eng
NIDN. 0402037401

HALAMAN PERUNTUKAN

Puji syukur kami panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya. Shalawat dan salam kami haturkan kepada Nabi Muhammad SAW. Dengan penuh rasa syukur, penulis dapat menyelesaikan skripsi ini yang berjudul “IMPLEMENTASI SIAMESE NEURAL NETWORK UNTUK DETEKSI KEMIRIPAN KODE : STUDI KASUS PENILAIAN TUGAS CODING MAHASISWA NUSA PUTRA” dengan baik.

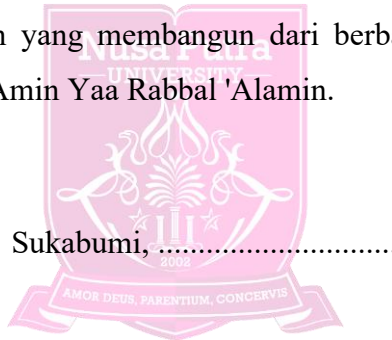
Penulis menyadari bahwa penulisan skripsi ini masih jauh dari kata sempurna, baik dari segi penyajian, permasalahan, maupun pemecahan masalah. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang konstruktif untuk perbaikan di masa mendatang. Semoga skripsi ini dapat bermanfaat, khususnya bagi penulis dan umumnya bagi semua pihak yang membutuhkan.

Selama mengerjakan skripsi ini, penulis tidak lepas dari bantuan dan dorongan dari berbagai pihak, sehingga skripsi ini dapat terselesaikan. Pada kesempatan ini, penulis ingin mengucapkan terima kasih yang setulus-tulusnya kepada:

1. Bapak Dr. H. Kurniawan, ST., M.Si., MM, selaku Rektor Universitas Nusa Putra.
2. Bapak Ir.H. Paikun, S.T.,IPM., Asean Eng, selaku Dekan Fakultas Univeristas Nusa Putra.
3. Bapak Ir.Somantri, ST., M.Kom, Selaku ketua program studi Teknik Informatika.
4. Bapak Nugraha, M.Kom, selaku Dosen Pembimbing 1 Universitas Nusa Putra, atas segala bimbingan, arahan serta saran yang diberikan kepada kami sehingga skripsi ini dapat diselesaikan dengan baik.
5. Ibu Ivana Lucia Kharisma, M.Kom, selaku dosen pembimbing 2 skripsi atas segala bimbingan, arahan serta saran yang diberikan kepada kami sehingga skripsi ini dapat diselesaikan dengan baik.
6. Ibu Anggun Fergina, M.Kom, selaku Ketua Penguji yang telah memberikan kritik dan saran yang membangun

7. Bapak Alun Sujjada, S.Kom, M.T, selaku pembimbing akademik kami, yang selalu memberikan bimbingan, arahan, dan dukungan selama proses perkuliahan.
8. Seluruh dosen/staff pengajar Fakultas Komputer Teknik dan desain program studi Teknik Informatika yang telah memberikan ilmu pengetahuan yang tak ternilai selama kami menempuh pendidikan di Universitas Nusa Putra.
9. Kepada seluruh keluarga, orang tua, kakak, adik, teman, anak, dan pasangan kami yang selalu memberikan dukungan penuh agar kami bisa menyelesaikan skripsi ini. Dorongan dan motivasi dari mereka sangat berarti, dan menjadi sumber semangat yang tak ternilai selama proses penulisan.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu kritik dan saran yang membangun dari berbagai pihak sangat kami harapkan demi perbaikan. Amin Yaa Rabbal 'Alamin.



Sukabumi,

Penulis I

Penulis II

Deni Setiawan

Ilham Nurabduljabbar Komara

ABSTRAK

Plagiarisme dalam tugas pemrograman menjadi tantangan signifikan dalam dunia akademik, terutama dalam proses penilaian yang adil dan objektif. Penelitian ini bertujuan untuk mengembangkan sistem deteksi kemiripan kode menggunakan arsitektur *Siamese Neural Network (SNN)* yang diintegrasikan dengan model *CodeBERT* untuk memahami kesamaan semantik antar potongan kode *Python*. *Dataset* diperoleh dari tugas mahasiswa Universitas Nusa Putra dan melalui tahap preprocessing seperti *augmentasi*, *embedding*, *labeling*, dan *balancing*. Model *SNN* dilatih untuk membedakan pasangan kode yang mirip dan tidak mirip dengan evaluasi menggunakan metrik akurasi, *presisi*, *recall*, dan *F1-score*. Hasil pelatihan menunjukkan bahwa sistem mampu mendeteksi kemiripan kode secara akurat. Sistem ini kemudian diimplementasikan dalam bentuk aplikasi *web* berbasis *Flask* dan *Firestore Authentication*, yang memungkinkan dosen untuk mengunggah, memeriksa, dan memantau hasil deteksi plagiarisme secara *real-time*. Penelitian ini diharapkan dapat membantu meningkatkan kualitas penilaian akademik dan mencegah terjadinya plagiarisme dalam tugas pemrograman.

Kata Kunci: *Siamese Neural Network*, *CodeBERT*, Deteksi Plagiarisme, Kemiripan Kode, Tugas Mahasiswa, *Deep Learning*, *Python*

ABSTRACT

Plagiarism in programming assignments poses a significant challenge in academic environments, especially in ensuring fair and objective assessment. This study aims to develop a code similarity detection system using a Siamese Neural Network (SNN) architecture integrated with the CodeBERT model to capture semantic similarities between Python code snippets. The dataset was collected from student assignments at Nusa Putra University and underwent preprocessing stages such as augmentation, embedding, labeling, and balancing. The SNN model was trained to distinguish between similar and dissimilar code pairs and evaluated using accuracy, precision, recall, and F1-score metrics. The training results indicate that the system can accurately detect code similarity. The system is implemented as a web application using Flask and Firebase Authentication, allowing lecturers to upload, examine, and monitor plagiarism detection results in real time. This research is expected to enhance academic assessment quality and help prevent plagiarism in programming assignments.

Keywords: Siamese Neural Network, CodeBERT, Plagiarism Detection, Code Similarity, Student Assignments, Deep Learning, Python.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “Implementasi Siamese Neural Network Untuk Deteksi Kemiripan Kode : Studi Kasus Penilaian Tugas Koding Mahasiswa Nusa Putra” ini dengan baik dan lancar. Shalawat serta salam senantiasa tercurah kepada junjungan kita Nabi Muhammad SAW, yang telah membawa umat manusia menuju jalan ilmu pengetahuan dan kebenaran.

Penyusunan skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika. Dalam proses penyusunan ini, penulis menyadari sepenuhnya bahwa keberhasilan dalam menyelesaikan tugas akhir ini tidak terlepas dari dukungan, bimbingan, dan bantuan dari berbagai pihak. Untuk itu, penulis menyampaikan rasa terima kasih dan penghargaan yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan baik secara moral maupun material.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, segala bentuk kritik dan saran yang membangun dari berbagai pihak sangat penulis harapkan sebagai bahan perbaikan di masa mendatang. Semoga karya sederhana ini dapat memberikan manfaat, baik bagi penulis sendiri maupun bagi pembaca serta pengembangan ilmu pengetahuan, khususnya di bidang Teknik Informatika. Aamiin yaa Rabbal ‘Aalamiin

Sukabumi,

Penulis 1

Penulis II

Deni Setiawan

Ilham Nurabduljabbar Komara

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas akademik UNIVERSITAS NUSA PUTRA, kami yang bertanda tangan dibawah ini :

Nama : Deni Setiawan
NIM : 20210040172
Nama : Ilham Nurabduljabbar Komara
NIM : 20210040011
Program Studi : Teknik Informatika
Jenis Karya : Tugas Akhir

Demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Nusa Putra **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

“IMPLEMENTASI SIAMESE NEURAL NETWORK UNTUK DETEKSI KEMIRIPAN KODE : STUDI KASUS PENILAIAN TUGAS CODING MAHASISWA NUSA PUTRA ”, beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Nusa Putra berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di:.....

Pada tanggal:.....

Yang Menyatakan,

Deni Setiawan

Ilham Nurabduljabbar Komara

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN PENULIS	ii
PENGESAHAN SKRIPSI.....	iii
HALAMAN PERUNTUKAN	iv
ABSTRAK	vi
<i>ABSTRACT</i>	vii
KATA PENGANTAR.....	viii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penelitian	5
BAB II TINJAUAN PUSTAKA.....	Error! Bookmark not defined.
2.1 Penelitian terkait.....	Error! Bookmark not defined.
2.2 Landasan Teori.....	Error! Bookmark not defined.
2.2.1 <i>Deep Learning</i>	Error! Bookmark not defined.
2.2.2 <i>CodeBERT</i>	Error! Bookmark not defined.
2.2.3 <i>Siamese Neural Network (SNN)</i>	Error! Bookmark not defined.
2.2.4 <i>Metric Learning</i>	Error! Bookmark not defined.
2.2.5 <i>Preprocessing Data</i>	Error! Bookmark not defined.
2.2.6 Metode Pengujian.....	Error! Bookmark not defined.
2.3 Kerangka Berfikir.....	Error! Bookmark not defined.
BAB III METODOLOGI PENELITIAN	Error! Bookmark not defined.
3.1 Metode Penelitian.....	Error! Bookmark not defined.

3.2	Tahapan Penelitian	Error! Bookmark not defined.
3.2.1	Studi Literatur	Error! Bookmark not defined.
3.2.2	Pengumpulan Data	Error! Bookmark not defined.
3.2.3	Pra-pemrosesan Data (<i>Data Preprocessing</i>)	Error! Bookmark not defined.
BAB IV HASIL DAN PEMBAHASAN.....		Error! Bookmark not defined.
4.1	Dataset.....	Error! Bookmark not defined.
4.2	<i>Preprocessing</i> Data Dataset	Error! Bookmark not defined.
4.3	<i>Pemodelan</i>	Error! Bookmark not defined.
4.4	Evalusi Model	Error! Bookmark not defined.
4.5	Implementasi Sitem <i>Website</i>	Error! Bookmark not defined.
4.6	Evaluasi Sistem Website	Error! Bookmark not defined.
BAB V PENUTUP.....		6
5.1	Kesimpulan	6
5.2	Saran.....	7
DAFTAR PUSTAKA.....		8



DAFTAR GAMBAR

Gambar 2. 1 <i>Arsitektur Deep Learning</i> [22]	Error! Bookmark not defined.
Gambar 2. 2 <i>Arsitektur CodeBERT</i> [29]	Error! Bookmark not defined.
Gambar 2. 3 <i>Arsitektur Siamese Neural Network</i> ..	Error! Bookmark not defined.
Gambar 2. 4 Kerangka Berfikir	Error! Bookmark not defined.
Gambar 3. 1 Tahapan Penelitian	Error! Bookmark not defined.
Gambar 3. 2 Tahapan Pra-pemrosesan data	Error! Bookmark not defined.
Gambar 3. 3 Tahapan Model <i>SDLC Waterfall</i>	Error! Bookmark not defined.
Gambar 3. 4 Arsitektur sistem	Error! Bookmark not defined.
Gambar 3. 5 Usecase Diagram	Error! Bookmark not defined.
Gambar 3. 6 <i>User Activity Login</i>	Error! Bookmark not defined.
Gambar 3. 7 <i>User Activity Registrasi</i>	Error! Bookmark not defined.
Gambar 3. 8 <i>User Activity</i> proses deteksi plagiarisme kode	Error! Bookmark not defined.
defined.	
Gambar 3. 9 <i>User Activity</i> Proses <i>Upload File</i>	Error! Bookmark not defined.
Gambar 3. 10 Proses Melihat Detail Perbandingan Kode	Error! Bookmark not defined.
defined.	
Gambar 3. 11 <i>User Activity Logout</i>	Error! Bookmark not defined.
Gambar 3. 12 <i>Wireframe Landing page</i>	Error! Bookmark not defined.
Gambar 3. 14 <i>Wireframe</i> Registrasi <i>page</i>	Error! Bookmark not defined.
Gambar 3. 15 <i>Wireframe Upload page</i>	Error! Bookmark not defined.
Gambar 3. 16 <i>Wireframe</i> monitoring <i>page</i>	Error! Bookmark not defined.
Gambar 4. 1 <i>Histogram</i> Hasil <i>Embbeding</i>	Error! Bookmark not defined.
Gambar 4. 2 Grafik sebelum <i>Balancing</i>	Error! Bookmark not defined.
Gambar 4. 3 Data setelah di <i>Balancing</i>	Error! Bookmark not defined.
Gambar 4. 4 Gambar <i>Output Training Model</i>	Error! Bookmark not defined.
Gambar 4. 5 <i>Confusion Matrix</i> Data Validasi Dan Uji	Error! Bookmark not defined.
defined.	
Gambar 4. 6 Grafik Akurasi <i>Loss</i> Pelatihan dan Validasi	Error! Bookmark not defined.
defined.	
Gambar 4. 7 <i>Landing page</i>	Error! Bookmark not defined.

Gambar 4. 8 Halaman *Login***Error! Bookmark not defined.**
 Gambar 4. 9 Halaman Registrasi**Error! Bookmark not defined.**
 Gambar 4. 10 Halaman *Upload File***Error! Bookmark not defined.**
 Gambar 4. 11 Tampilan Halaman monitoring dan hasil pemeriksaan..... **Error!**
Bookmark not defined.

DAFTAR TABEL

Tabel 2. 1 Perbandingan dengan penelitian terdahulu**Error! Bookmark not defined.**
 Tabel 3. 1 Confusion Matrix**Error! Bookmark not defined.**
 Tabel 3. 2 Pengujian sistem**Error! Bookmark not defined.**
 Tabel 4. 1 Jumlah *Dataset Kode Python*.....**Error! Bookmark not defined.**
 Tabel 4. 2 lingkungan eksperimen.....**Error! Bookmark not defined.**
 Tabel 4. 3 skema ekstraksi**Error! Bookmark not defined.**
 Tabel 4. 4 skema ekstraksi**Error! Bookmark not defined.**
 Tabel 4. 5 data hasil labeling.....**Error! Bookmark not defined.**
 Tabel 4. 6 hasil dari proses *pairing* dan labeling ...**Error! Bookmark not defined.**
 Tabel 4. 7 Dataset final**Error! Bookmark not defined.**
 Tabel 4. 8 isi file dari seluruh komponen.....**Error! Bookmark not defined.**
 Tabel 4. 9 *Output* akhir**Error! Bookmark not defined.**
 Tabel 4. 10 Tabel Pembagian Dataset.....**Error! Bookmark not defined.**
 Tabel 4. 11 Tabel Parameter Pelatihan**Error! Bookmark not defined.**
 Tabel 4. 12 Tabel Validasi akurasi.....**Error! Bookmark not defined.**
 Tabel 4. 13 Tabel Hasil Evaluasi Model pada Data uji**Error! Bookmark not defined.**
defined.
 Tabel 4. 14 Tabel Teknologi yang digunakan.....**Error! Bookmark not defined.**
 Tabel 4. 15 Hasil BlackBox Testing**Error! Bookmark not defined.**

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digital saat ini, keterampilan pemrograman menjadi salah satu kemampuan yang sangat penting di berbagai bidang seperti teknologi informasi, rekayasa perangkat lunak dan ilmu computer [1]. Banyak perguruan tinggi yang menawarkan program studi di bidang ini, salah satunya mata kuliah dasar yang menjadi pondasi penting adalah Dasar Pemrograman. Mata kuliah ini bertujuan untuk membekali mahasiswa dengan kemampuan logika dan struktur dasar pemrograman, seperti penggunaan variabel, percabangan, perulangan dan fungsi. Penilaian pada mata kuliah ini umumnya berbentuk tugas-tugas berupa kode program yang harus dikerjakan dan dikumpulkan oleh mahasiswa.

Salah satu masalah utama yang dihadapi dosen adalah mendeteksi kemiripan kode dalam mata kuliah Dasar Pemrograman. Kemiripan kode dapat menjadi indikasi plagiarisme yang merupakan masalah serius dalam dunia akademis [2]. Plagiarisme tidak hanya merugikan integritas akademik, tetapi juga menghambat proses pembelajaran mahasiswa. Plagiarisme dalam pemrograman dapat mengurangi kemampuan mahasiswa untuk belajar dan memahami konsep-konsep dasar pemrograman [3] karena sering kali terjadi saat mahasiswa mengerjakan tugas di mana banyak ditemukan kemiripan dalam *Source Code* yang dikumpulkan akibat menyalin hasil penelitian atau merangkum karya orang lain tanpa menyebutkan sumbernya [4]. Sehingga hal ini bisa menyebabkan adanya ketergantungan terhadap orang lain dalam mengerjakan tugas dan tidak dapat secara mandiri mengerjakan tugas yang diberikan oleh dosen khususnya dalam mata kuliah pemrograman [5]. Oleh karena itu, penting untuk mengembangkan sistem yang dapat secara otomatis mendeteksi kemiripan kode sehingga dosen dapat memberikan penilaian yang lebih adil dan objektif.

Sebelum berkembangnya teknologi *deep learning*, deteksi plagiarisme dan duplikasi dalam *source code* mengandalkan alat pendeteksi kemiripan otomatis [6], beberapa di antaranya seperti *JPlag* [7] dan *Algoritma Winnowing* [5]. Meskipun efektif untuk mendeteksi kemiripan sintaksis, metode-metode tradisional ini

memiliki keterbatasan dalam menangkap kemiripan semantik, terutama ketika kode mengalami transformasi seperti perubahan nama variabel atau restrukturisasi [8].

Seiring dengan kemajuan teknologi *deep learning*, paradigma deteksi kemiripan kode mengalami transformasi signifikan. Arsitektur *deep learning* mampu secara otomatis mempelajari representasi hierarkis dan abstrak dari data kode, memungkinkan pemahaman kemiripan pada level semantik yang lebih mendalam [9]. *Siamese Neural Networks* merupakan salah satu arsitektur *deep learning* yang dirancang khusus untuk pembelajaran metrik (*metric learning*) dan tugas perbandingan kemiripan [10]. Arsitektur ini terdiri dari dua sub-jaringan neural yang identik dengan berbagi bobot (*weight sharing*) memungkinkan model mempelajari representasi *vektor* optimal di mana objek-objek yang mirip memiliki jarak yang dekat dalam ruang fitur [11].

Dalam konteks deteksi kemiripan kode, model ini dapat digunakan untuk membandingkan dua potongan kode dan menghasilkan nilai kemiripan berdasarkan representasi fitur yang dipelajari [12]. Penelitian ini menggunakan dataset berbentuk teks yang berisi kode program *Python* yang kemudian diproses menggunakan teknik *preprocessing* seperti *tokenisasi* dan *normalisasi* untuk mengoptimalkan kinerja *model* [13]. Setelah proses pelatihan dan evaluasi selesai, model kemudian diintegrasikan ke dalam *back-end* menggunakan *framework* seperti *Flask* untuk memudahkan pengguna dalam mengakses dan menggunakan sistem secara *real-time* [14].

Selain itu, penelitian ini turut berkontribusi dalam mengembangkan metode penilaian yang lebih objektif melalui otomatisasi deteksi kemiripan dalam tugas kode program. Dengan adanya sistem ini, dosen dan mahasiswa dapat memverifikasi setiap tugas secara menyeluruh guna mencegah terjadinya plagiarisme akademik. Fitur utama dari sistem ini meliputi deteksi kemiripan kode secara presisi, visualisasi hasil perbandingan, serta laporan evaluasi otomatis. Manfaat yang diperoleh pengguna meliputi efisiensi waktu dalam proses penilaian, peningkatan transparansi dalam evaluasi, serta jaminan bahwa penilaian dilakukan berdasarkan kualitas dan orisinalitas kode yang ditulis oleh mahasiswa.

Secara keseluruhan, penelitian ini bertujuan untuk memberikan solusi yang inovatif dalam menghadapi tantangan penilaian tugas *coding*. Dengan

memanfaatkan teknologi terkini seperti *Siamese Neural Networks*, sistem yang dikembangkan dapat memberikan manfaat yang signifikan bagi dunia pendidikan, khususnya dalam penilaian tugas *coding* mahasiswa.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, permasalahan dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana merancang arsitektur *Siamese Neural Networks* yang optimal untuk mendeteksi kemiripan kode dalam tugas *coding* mahasiswa?
2. Bagaimana tingkat akurasi sistem yang dikembangkan dalam mendeteksi kemiripan kode dibandingkan dengan penilaian manual yang dilakukan oleh dosen?
3. Bagaimana implementasi sistem berbasis *web* yang dapat memfasilitasi proses deteksi kemiripan kode secara efisien dan *user-friendly*?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan fokus, maka ruang lingkup permasalahan dibatasi sebagai berikut:

1. Penelitian ini hanya fokus pada deteksi kemiripan kode dalam bahasa pemrograman *Python* untuk mata kuliah Dasar Pemrograman. Kode dalam bahasa pemrograman lain tidak akan dianalisis dalam penelitian ini.
2. *Dataset* yang digunakan dalam penelitian ini adalah tugas *coding* mahasiswa dari Universitas Nusa Putra yang telah dikumpulkan selama satu semester. Tugas yang tidak lengkap atau tidak sesuai dengan kriteria akan dieksklusi dari analisis.
3. Penelitian ini menggunakan arsitektur *Siamese Neural Networks* dengan sub-jaringan berbasis *LSTM (Long Short-Term Memory)* untuk pemrosesan sekuensial kode. Model lain seperti *CNN* atau *Transformer* tidak akan dipertimbangkan dalam penelitian ini.
4. Evaluasi sistem difokuskan pada metrik akurasi, *presisi*, *recall* dan *F1-score*. Analisis kompleksitas waktu komputasi dan penggunaan memori tidak menjadi fokus utama penelitian.

5. Implementasi sistem berbasis web menggunakan framework *Flask* untuk back-end dan teknologi *web* standar (*HTML*, *CSS*, *JavaScript*) untuk *front-end*. Integrasi dengan *platform* pembelajaran lain tidak termasuk dalam ruang lingkup penelitian.

1.4 Tujuan Penelitian

Selaras dengan rumusan masalah yang telah ditetapkan, tujuan dari penelitian ini adalah:

1. Merancang dan mengimplementasikan arsitektur *Siamese Neural Networks* yang dapat mengekstraksi fitur semantik dari kode *Python* secara efektif untuk tugas deteksi kemiripan.
2. Mengukur dan menganalisis performa sistem dalam mendeteksi kemiripan kode melalui evaluasi komprehensif menggunakan metrik akurasi, *presisi*, *recall*, dan *F1-score*.
3. Mengembangkan aplikasi *web* yang terintegrasi dengan *model Siamese Neural Networks* untuk memfasilitasi proses deteksi kemiripan kode secara otomatis dan menyediakan antarmuka yang intuitif bagi pengguna.
4. Menyediakan sistem penilaian objektif yang dapat membantu dosen dalam mengevaluasi orisinalitas tugas coding mahasiswa dan mencegah praktik plagiarisme akademik.

1.5 Manfaat Penelitian

Adapun untuk beberapa manfaat dari penelitian ini adalah sebagai berikut :

1. Penelitian ini akan memberikan kemudahan bagi dosen dalam menilai tugas coding mahasiswa. Dengan sistem deteksi kemiripan kode, dosen tidak perlu lagi memeriksa setiap tugas secara manual, sehingga dapat menghemat waktu dan tenaga.
2. Dengan menggunakan teknologi seperti *Siamese Neural Networks*, sistem ini dapat memberikan penilaian yang lebih tepat. Hal ini akan mengurangi kemungkinan kesalahan dalam menilai kemiripan kode sehingga mahasiswa mendapatkan penilaian yang lebih adil.
3. Sistem ini juga berfungsi untuk mencegah plagiarisme di kalangan mahasiswa. Dengan mendeteksi kemiripan kode secara otomatis,

mahasiswa akan lebih termotivasi untuk menulis kode mereka sendiri dan memahami konsep pemrograman dengan lebih baik.

4. Penelitian ini dapat memberikan wawasan dan dasar bagi penelitian selanjutnya di bidang deteksi kemiripan kode. Hasil dan temuan dari penelitian ini dapat digunakan oleh peneliti lain untuk mengembangkan metode yang lebih baik dalam mendeteksi kemiripan kode.

1.6 Sistematika Penelitian

Memberikan gambaran secara garis besar, dalam hal ini dijelaskan isi dari masing-masing bab dari tugas akhir ini. Sistematika penulisan dalam pembuatan laporan ini sebagai berikut :

1. BAB I PENDAHULUAN

Bab ini berisikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, dan sistematika penelitian tentang Pembuatan Implementasi *Siamese Neural Networks* untuk deteksi kemiripan kode: Studi Kasus Penilaian Tugas *Coding* Mahasiswa Nusa Putra.

2. BAB II LANDASAN TEORI

Pada bab ini berisikan tentang dasar teori penjelasan, penelitian terdahulu dan deskripsi kegiatan penelitian.

3. BAB III METODE PENELITIAN

Bab ini berisikan tentang metode yang digunakan yaitu metode *Siamese Network* dan cara kerja pendeteksian kemiripan pada kode.

4. BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini membahas tentang output yang dihasilkan dalam penelitian dan dibahas secara detail.

5. BAB V PENUTUP

Dalam bab ini membahas tentang kesimpulan dari isi semua penelitian skripsi ini dan diakhiri dengan penutup.

BAB V

PENUTUP

5.1 Kesimpulan

Penelitian ini berhasil merancang dan mengimplementasikan sebuah sistem pendeteksi plagiarisme kode program berbasis web menggunakan arsitektur *Siamese Neural Network* dan model *pre-trained CodeBERT*. Berdasarkan tahapan penelitian yang dilakukan, diperoleh beberapa kesimpulan utama:

1. Pengumpulan dan Preprocessing Data

Dataset diperoleh dari tugas-tugas pemrograman *Python* milik 15 mahasiswa dengan total 165 file asli. Setelah dilakukan augmentasi sebanyak 5 kali per file, jumlah total file menjadi 990. File ini kemudian dikonversi menjadi vektor *embedding* berdimensi 768 menggunakan metode CLS Token dari *CodeBERT* dan dinormalisasi menggunakan *L2 normalization*.

2. Pemodelan dan Pelatihan

Model Siamese Neural Network dilatih dengan 3300 pasangan data (50% mirip, 50% tidak mirip). Dataset dibagi ke dalam 60% data latih, 20% data validasi, dan 20% data uji. Model dilatih selama beberapa *epoch* dengan teknik *early stopping* dan menghasilkan akurasi tinggi pada data uji sebesar 90.45%.

3. Evaluasi Sistem

Evaluasi dilakukan dengan metrik klasifikasi yang menunjukkan performa sangat baik, dengan Precision 84.20%, Recall 99.39%, dan F1-Score 91.16%. Hal ini membuktikan bahwa sistem mampu membedakan kode yang mirip maupun tidak mirip secara efektif.

4. Implementasi Web dan Pengujian

Sistem berhasil diintegrasikan ke dalam aplikasi web sederhana menggunakan *Flask* untuk *backend* dan *HTML/JS* untuk *frontend*. Pengujian menggunakan metode *black-box* menunjukkan semua fungsi berjalan sesuai ekspektasi, termasuk *login*, *upload file*, deteksi plagiarisme, serta tampilan hasil.

5.2 Saran

Beberapa saran untuk pengembangan lebih lanjut dari sistem ini antara lain:

1. Perluasan Deteksi ke Bahasa Lain

Sistem saat ini hanya mendukung *file Python*. Pengembangan selanjutnya bisa mencakup bahasa pemrograman lain seperti *C++*, *Java*, atau *JavaScript*.

2. Integrasi dengan LMS

Sistem dapat diperluas agar terhubung langsung dengan sistem pembelajaran daring (LMS) seperti *Edlink* atau *Moodle*, sehingga proses upload dan pemeriksaan dapat dilakukan secara otomatis dan terintegrasi.

3. Penambahan Penjelasan Hasil ke Mahasiswa

Untuk mendukung transparansi dan pembelajaran, sistem bisa menampilkan highlight perbedaan kode atau skor kemiripan per baris kode.



DAFTAR PUSTAKA

- [1] Febby Wilyani, Qonaah Nuryan Arif, and Fitri Aslimar, “Pengenalan Dasar Pemrograman Python Dengan Google Colaboratory,” *Jurnal Pelayanan dan Pengabdian Masyarakat Indonesia*, vol. 3, no. 1, pp. 08–14, 2024, doi: 10.55606/jppmi.v3i1.1087.
- [2] O. Karnalim, Simon, and W. Chivers, “Layered similarity detection for programming plagiarism and collusion on weekly assessments,” *Computer Applications in Engineering Education*, vol. 30, no. 6, pp. 1739–1752, 2022, doi: 10.1002/cae.22553.
- [3] D. A. Khairani and Z. Zainarti, “Tinjauan Mendalam Tentang Plagiarisme: Pelanggaran Etika dalam Dunia Akademik dan Profesional,” 2025.
- [4] I Gede Ariawan Eka Putraa and I Wayan Supriana, “Deteksi Plagiarisme Source Code Tugas Mahasiswa Menggunakan Algoritma Cosine Similarity dan Pembobotan TF-IDF,” 2022.
- [5] M. S. Ramli, S. Cokrowibowo, and M. F. Rustan, “Uji Plagiarism pada Tugas Mahasiswa Menggunakan Algoritma Winnowing,” *Journal of Applied Computer Science and Technology*, vol. 2, no. 2, pp. 108–112, Dec. 2021, doi: 10.52158/jacost.v2i2.177.
- [6] G. Lee, J. Kim, M. S. Choi, R. Y. Jang, and R. Lee, “Review of Code Similarity and Plagiarism Detection Research Studies,” *Applied Sciences (Switzerland)*, vol. 13, no. 20, Oct. 2023, doi: 10.3390/app132011358.
- [7] T. Sağlam and L. Schmid, “Evaluating Software Plagiarism Detection in the Age of AI: Automated Obfuscation and Lessons for Academic Integrity,” May 2025, [Online]. Available: <http://arxiv.org/abs/2505.20158>
- [8] E. N. Akimova *et al.*, “A Survey on Software Defect Prediction Using Deep Learning,” *Numerical Analysis and Scientific Computing*, vol. 11, no. 9, May 2021, doi: <https://doi.org/10.3390/math9111180>.
- [9] A. Raup, W. Ridwan, Y. Khoeriyah, S. Supiana, and Q. Y. Zaqiah, “Deep Learning dan Penerapannya dalam Pembelajaran,” *JIIIP - Jurnal Ilmiah Ilmu Pendidikan*, vol. 5, no. 9, pp. 3258–3267, 2022, doi: 10.54371/jiip.v5i9.805.
- [10] A. Foroozandeh, A. A. Hemmat, and H. Rabbani, “Offline Handwritten Signature Verification and Recognition Based on Deep Transfer Learning,” *IEEE*, pp. 7683–7690, Sep. 2020, doi: <https://doi.org/10.1109/MVIP49855.2020.9187481>.

- [11] P. R. Arum, M. Dzeaulfath, and A. O. Winarta, "One-Shot Learning Menggunakan Siamese Neural Network Untuk Pendeteksian Wajah," in *Prosiding Seminar Nasional Sains Data*, Universitas Muhammadiyah Semarang, Nov. 2023, pp. 290–295. doi: <https://doi.org/10.33005/senada.v3i1.125>.
- [12] H. Zhang and K. Sakurai, "A Survey of Software Clone Detection from Security Perspective," 2021, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2021.3065872.
- [13] O. Karnalim, Simon, and W. Chivers, "Preprocessing for Source Code Similarity Detection in Introductory Programming," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Nov. 2020. doi: 10.1145/3428029.3428065.
- [14] M. Grinberg, "The New and Improved Flask Mega-Tutorial (2024 Edition)," 2025. [Online]. Available: <https://github.com/miguelgrinberg/microblog/tree/v0.4>
- [15] N. Awale, M. Pandey, A. Dulal, and B. Timsina, "Plagiarism Detection in Programming Assignments using Machine Learning," *Journal of Artificial Intelligence and Capsule Networks*, vol. 2, no. 3, pp. 177–184, 2020, doi: 10.36548/jaicn.2020.3.005.
- [16] E. Araujo-pino and G. Fuentes-pineda, "Siamese Network applied to Authorship Verification," *Clef 2020*, no. September, pp. 22–25, 2020.
- [17] M. Fisichella, "Siamese coding network and pair similarity prediction for near-duplicate image detection," *Int J Multimed Inf Retr*, vol. 11, no. 2, pp. 159–170, 2022, doi: 10.1007/s13735-022-00233-w.
- [18] O. Karnalim and G. Kurniawati, "Programming style on source code plagiarism and collusion detection," *International Journal of Computing*, vol. 19, no. 1, pp. 27–38, 2020, doi: 10.47839/ijc.19.1.1690.
- [19] Z. Tian, Q. Wang, C. Gao, L. Chen, and D. Wu, "Plagiarism Detection of Multi-Threaded Programs via Siamese Neural Networks," *IEEE Access*, vol. 8, pp. 160802–160814, 2020, doi: 10.1109/ACCESS.2020.3021184.
- [20] T. Rüz and C. Beisbart, "The Importance of Understanding Deep Learning," *Erkenntnis*, vol. 89, no. 5, pp. 1823–1840, Jun. 2024, doi: 10.1007/s10670-022-00605-y.
- [21] B. Sistaninejhad, H. Rasi, and P. Nayeri, "A Review Paper about Deep Learning for Medical Image Analysis," 2023, *Hindawi Limited*. doi: 10.1155/2023/7091301.

- [22] J. Nurhakiki *et al.*, “Studi Kepustakaan: Pengenalan 4 Algoritma Pada Pembelajaran Deep Learning Beserta Implikasinya,” *Jurnal Pendidikan Berkarakter*, no. 1, pp. 270–281, 2024, doi: 10.51903/pendekar.v2i1.598.
- [23] M. Rizki, S. Basuki, and Y. Azhar, “Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory Untuk Prediksi Curah Hujan Kota Malang,” *REPOSITOR*, vol. 2, no. 3, pp. 331–338, 2020.
- [24] Z. Feng *et al.*, “CodeBERT: A Pre-Trained Model for Programming and Natural Languages,” Sep. 2020, [Online]. Available: <http://arxiv.org/abs/2002.08155>
- [25] W. Wang, G. Li, B. Ma, X. Xia, and Z. Jin, “Detecting Code Clones with Graph Neural Network and Flow-Augmented Abstract Syntax Tree,” Feb. 2020, [Online]. Available: <http://arxiv.org/abs/2002.08653>
- [26] H. Husain, H.-H. Wu, T. Gazit, M. Allamanis, and M. Brockschmidt, “CodeSearchNet Challenge: Evaluating the State of Semantic Code Search,” Jun. 2020, [Online]. Available: <http://arxiv.org/abs/1909.09436>
- [27] Z. Feng *et al.*, “CodeBERT: A Pre-Trained Model for Programming and Natural Languages,” Sep. 2020, [Online]. Available: <http://arxiv.org/abs/2002.08155>
- [28] W. U. Ahmad, S. Chakraborty, B. Ray, and K.-W. Chang, “Unified Pre-training for Program Understanding and Generation,” Apr. 2021, [Online]. Available: <http://arxiv.org/abs/2103.06333>
- [29] F. J. Peña, A. Luis Gonzalez, S. Pashami, A. Al-Shishtawy, and A. H. Payberah, “SIAMBERT: Siamese Bert-based Code Search,” 2022.
- [30] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, “MagFace: A Universal Representation for Face Recognition and Quality Assessment,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, Ithaca: IEEE, Jul. 2021, pp. 14225–14234.
- [31] N. Mehrotra, N. Agarwal, P. Gupta, S. Anand, D. Lo, and R. Purandare, “Modeling Functional Similarity in Source Code with Graph-Based Siamese Networks,” Nov. 2020, [Online]. Available: <http://arxiv.org/abs/2011.11228>
- [32] J. Li, Y. Q. Zhou, and Q. Y. Zhang, “Metric networks for enhanced perception of non-local semantic information,” *Front Neurobot*, vol. 17, 2023, doi: 10.3389/fnbot.2023.1234129.

- [33] P. Khosla *et al.*, “Supervised Contrastive Learning,” in *Neural Information Processing Systems*, Vancouver, Canada, Mar. 2021, pp. 1–23. [Online]. Available: <http://arxiv.org/abs/2004.11362>
- [34] M. Kim, A. K. Jain, and X. Liu, “AdaFace: Quality Adaptive Margin for Face Recognition,” in *IEEE*, Michigan, Feb. 2023, pp. 18750–18759. [Online]. Available: <http://arxiv.org/abs/2204.00964>
- [35] C. Yan, G. Pang, X. Bai, J. Zhou, and L. Gu, “Beyond Triplet Loss: Person Re-identification with Fine-grained Difference-aware Pairwise Loss,” Sep. 2020, [Online]. Available: <http://arxiv.org/abs/2009.10295>
- [36] D. Aganjan, M. Eisenbach, J. Wagner, D. Seichter, and H.-M. Gross, “Revisiting Loss Functions for Person Re-identification,” in *Int. Conf on Artificial Neural Networks (ICANN)*, Bratislava, Nov. 2021, p. 30. [Online]. Available: <http://arxiv.org/abs/1703.07737>
- [37] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, “Barlow Twins: Self-Supervised Learning via Redundancy Reduction,” in *Proceedings of Machine Learning Research*, Meila, Marina, and T. Zhang, Eds., PMLR, Jun. 2021, pp. 12310–12320. doi: <https://doi.org/10.48550/arXiv.2103.03230>.
- [38] F. Alghifari and D. Juardi, “Fauzan Alghifari Penerapan Data Mining Pada Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes,” 2021.
- [39] S. Dika Pratama and M. Novinarsyah Dadaprawira, “Pengujian Black Box Testing Pada Aplikasi Edu Digital Berbasis Website Menggunakan Metode Equivalence Dan Boundary Value,” *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 6, no. 2, pp. 560–569, 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsk/index>
- [40] F. Nurrisa and D. Hermina, “Pendekatan Kualitatif dalam Penelitian: Strategi, Tahapan, dan Analisis Data,” vol. 02, pp. 793–800, 2025.
- [41] D. Arya Pamungkas, H. Nurcahya Sumirat, S. Sahira, I. Lucia Kharisma, and A. Fergina, “Jurnal Rekayasa Teknologi Nusa Putra SISTEM INFORMASI PENDAFTARAN IZIN MASUK KAWASAN KONSERVASI (SIMAKSI) BERBASIS WEB PADA JALUR PENDAKIAN TAMAN NASIONAL GUNUNG HALIMUN SALAK,” vol. 10, no. 2, pp. 103–110, 2024, [Online]. Available: <https://rekayasa.nusaputra.ac.id/index>