UNIVERSITAS
GADJAH MADA
**Deteksi Code Smell Menggunakan Kombinasi Metrik Struktural dan Representasi Semantik Kode pada
Dataset yang Relevan dengan Industri**
Azhar Bagaskara, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR ISI

Halaman Judul **ii**

Halaman Pengesahan **iii**

Halaman Pernyataan **iv**

PRAKATA **v**

INTISARI **xiii**

ABSTRACT **xiv**

I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Tujuan Penelitian 4
1.4 Manfaat Penelitian 5
1.5 Batasan Penelitian 5

II TINJAUAN PUSTAKA **6**
2.1 Deteksi *Code Smell* Berbasis Metode Heuristik 6
2.2 Deteksi *Code Smell* Menggunakan *Machine Learning* 6
2.3 Penggunaan *Neural Code Embedding* Sebagai Fitur Semantik Pada Deteksi *Code Smell* 7
2.4 Pendekatan Kombinasi Fitur Struktural dan Semantik dalam Mendeteksi *Code Smell* 8
2.5 Tantangan dan Kontribusi 8

III LANDASAN TEORI **15**
3.1 *Code Smell* 15
3.1.1 *Feature envy* 15
3.1.2 *Long method* 16
3.1.3 *Blob* 18
3.1.4 *Data class* 20
3.2 Dataset Madeyski Lewowski Code Quest (MLCQ) 21

UNIVERSITAS
GADJAH MADA
**Deteksi Code Smell Menggunakan Kombinasi Metrik Struktural dan Representasi Semantik Kode pada
Dataset yang Relevan dengan Industri**
Azhar Bagaskara, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

3.3 Fitur Struktural . . . 22
3.3.1 Metrik kode . . . 22
3.4 Fitur Semantik . . . 25
3.4.1 code2vec . . . 25
3.4.2 CuBERT . . . 27
3.4.3 CodeBERT . . . 28
3.5 *Random Forests* . . . 30
3.6 XGBoost . . . 31
3.7 *Support Vector Machine* (SVM) . . . 33
3.8 *Principal Component Analysis* (PCA) . . . 35
3.9 *Synthetic Minority Over-sampling Technique* (SMOTE) . . . 37
3.10 Metrik Evaluasi . . . 37
3.10.1 *Precision* . . . 38
3.10.2 *Recall* . . . 38
3.10.3 *F1 score* . . . 38
3.10.4 *Matthews Correlation Coefficient* (MCC) . . . 39

IV METODE PENELITIAN 40
4.1 Deskripsi Penelitian . . . 40
4.2 Studi Literatur . . . 40
4.3 Pengumpulan Data . . . 40
4.4 Penyesuaian Dataset . . . 42
4.5 Ekstraksi Fitur . . . 44
4.5.1 Fitur struktural . . . 44
4.5.2 Fitur semantik . . . 46
4.6 *Data Preprocessing* . . . 48
4.6.1 Imputasi *missing value* . . . 48
4.6.2 Standarisasi . . . 49
4.6.3 *Dimensionality reduction* untuk fitur semantik . . . 49
4.6.4 *Oversampling* . . . 50
4.7 *Train Test Split* . . . 51
4.8 *Modeling* . . . 51
4.9 Evaluasi . . . 53

UNIVERSITAS
GADJAH MADA
**Deteksi Code Smell Menggunakan Kombinasi Metrik Struktural dan Representasi Semantik Kode pada
Dataset yang Relevan dengan Industri**
Azhar Bagaskara, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

V IMPLEMENTASI **54**
5.1 Perangkat Implementasi 54
5.2 Persiapan *Library* 54
5.3 Implementasi Pengumpulan Data 55
5.4 Implementasi Penyesuaian Dataset 56
5.5 Implementasi Ekstraksi Fitur Struktural 58
5.6 Implementasi Ekstraksi Fitur Semantik 60
5.6.1 Ekstraksi vektor *embedding* code2vec 60
5.6.2 Ekstraksi vektor *embedding* CuBERT 64
5.6.3 Ekstraksi vektor *embedding* CodeBERT 65
5.7 Implementasi *Data Preprocessing* 66
5.7.1 Prapemrosesan data skenario *baseline* 66
5.7.2 Prapemrosesan data skenario kombinasi fitur 67
5.8 Implementasi Pembangunan dan Pelatihan Model 68
5.8.1 Pembagian data *stratified hold-out* 68
5.8.2 Optimasi *hyperparameter* dengan Optuna 69
5.8.3 Pelatihan akhir dan evaluasi model 71

VI HASIL DAN PEMBAHASAN **72**
6.1 Hasil Pengumpulan Data 72
6.2 Hasil Penyesuaian Dataset 72
6.3 Hasil Evaluasi Model 74
6.3.1 Evaluasi model fitur struktural 74
6.3.2 Evaluasi model fitur semantik 75
6.3.3 Evaluasi model fitur kombinasi 77
6.4 Komparasi dan Analisis Hasil Evaluasi 80
6.4.1 Perbandingan antara penggunaan fitur tunggal dan fitur kombinasi 81
6.4.2 Pengaruh algoritma *machine learning* pada fitur kombinasi . . 83
6.4.3 Pengaruh jenis fitur semantik pada fitur kombinasi 85
6.4.4 Analisis kesepakatan klasifikasi menggunakan Cohen's Kappa pada model fitur kombinasi terbaik 87
6.4.5 Analisis *feature importance* pada model fitur kombinasi terbaik 88
6.5 Diskusi Perbandingan Hasil 91
6.5.1 Analisis efektivitas kombinasi fitur terhadap model *baseline* . 91

UNIVERSITAS
GADJAH MADA
**Deteksi Code Smell Menggunakan Kombinasi Metrik Struktural dan Representasi Semantik Kode pada
Dataset yang Relevan dengan Industri**
Azhar Bagaskara, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

6.5.2 Perbandingan hasil dengan penelitian terdahulu 95

VII KESIMPULAN DAN SARAN **97**

7.1 Kesimpulan 97

7.2 Saran 98