UNIVERSITAS
GADJAH MADA
Klasifikasi Gejala Depresi pada Teks Postingan Reddit dengan Pendekatan XGBoost dan Random Forest
Shevrilla Vilnafa Bilbina S., Dr. Sigit Priyanta, S.Si., M.Kom.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR ISI

HALAMAN PENGESAHAN....4
HALAMAN PERNYATAAN....4
HALAMAN PERSEMBAHAN....5
PRAKATA....6
DAFTAR ISI....8
DAFTAR GAMBAR....11
DAFTAR TABEL....13
ABSTRAK....14
BAB I PENDAHULUAN....16
1.1 Latar Belakang....16
1.2 Rumusan Masalah....18
1.3 Tujuan Penelitian....19
1.4 Manfaat Penelitian....19
BAB II TINJAUAN PUSTAKA....21
BAB III LANDASAN TEORI....27
3.1 Depresi....27
3.2 Feature Extraction....27
3.2.1 N-Grams....28
3.2.2 TF-IDF....28
3.2.3 Word2Vec....29
3.3 Machine Learning Algorithms....30
3.4 Random Forest....30
3.5 XGBoost....31
3.7 Hyper-parameter Tuning....32
3.8 Metrik Evaluation Model....33
3.8.1 Accuration....34
3.8.2 Precision....34
3.8.3 Recall....35
3.8.4 F-1 Score....35
BAB IV METODOLOGI....37
4.1 Deskripsi Umum Penelitian....37
4.2 Akuisisi Data....38
4.2.1 Text Processing....41
4.2.1.1 Cleaning....42
4.2.1.2 Stop Word Removals....42
4.2.1.3 Stemming/Lemmatization....43
4.2.2 Ekstraksi Fitur....43
4.2.2.1 Representasi dengan N-gram....44
4.2.2.2 Representasi dengan TF-IDF....45

UNIVERSITAS
GADJAH MADA
Klasifikasi Gejala Depresi pada Teks Postingan Reddit dengan Pendekatan XGBoost dan Random Forest
Shevrilla Vilnafa Bilbina S., Dr. Sigit Priyanta, S.Si., M.Kom.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

4.2.2.3 Representasi dengan Word2Vec....45
4.2.3 Modelling....46
4.2.4 Evaluation and Analysis....47
4.3 Implementasi....49
4.4 Strategi Evaluasi dan Pengujian Performa Metode....49
4.5 Penulisan Laporan....50
BAB V IMPLEMENTASI....51
5.1 Pengambilan Data dalam Postingan Reddit....51
5.2 Prapemprosesan Data....53
5.2.1 Normalisasi Data....53
5.2.2 Penghapusan Tanda Baca....54
5.2.3 Penghapusan Stopword....56
5.2.4 Implementasi Lemmatization....58
5.3 Ekstraksi Fitur N-gram....60
5.4 Visualisasi Wordcloud....62
5.5 Implementasi Word2Vec....64
5.5 Implementasi TF-IDF....67
5.6 Implementasi Ngram....68
5.7 Data Preparation....70
5.7.1 Data Preparation Ngram....70
5.7.2 Data Preparation Word2Vec....71
5.7.3 Data Preparation TF-IDF....72
5.8 Pembagian data latih dan data uji....73
5.9 Implementasi Random Forest dengan Fitur Ngram....74
5.10 Implementasi Random Forest dengan Fitur Word2Vec....76
5.11 Implementasi Random Forest dengan Fitur TF-IDF....77
5.12 Implementasi XGBoost dengan Fitur Ngram....78
5.13 Klasifikasi XGBoost dengan Fitur Word2Vec....80
5.14 Implementasi XGBoost dengan Fitur TF-IDF....82
5.15 Implementasi Hyperparameter pada Random Forest dan N-gram....84
5.16 Implementasi Hyperparameter pada Random Forest dan Word2Vec....86
5.17 Implementasi Hyperparameter pada Random Forest dan TF-IDF....87
5.18 Implementasi Hyperparameter pada XGBoost dan N-gram....89
5.19 Implementasi Hyperparameter pada XGBoost dan Word2Vec....91
5.20 Implementasi Hyperparameter pada XGBoost dan TF-IDF....93
5.21 Final Prediction....95
BAB VI HASIL DAN PEMBAHASAN....98
6.1 Hasil Implementasi Model....98
6.2 Confusion Matrix....105
6.2 ROC Curve dan Precision-Recall Curve....106
6.4 Comparative Statistical Analysis....109
6.5 Feature Importance....112
6.6 Learning Curve....113
6.7 Ablation Study....114

UNIVERSITAS
GADJAH MADA
Klasifikasi Gejala Depresi pada Teks Postingan Reddit dengan Pendekatan XGBoost dan Random Forest
Shevrilla Vilnafa Bilbina S., Dr. Sigit Priyanta, S.Si., M.Kom.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

6.8 Perbandingan Hasil Implementasi Model 115
BAB VII KESIMPULAN DAN SARAN 118
7.1 Kesimpulan 118
7.2 Saran 118
DAFTAR PUSTAKA 119