DAFTAR ISI

HALAMAN PENGESAHAN PEMBIMBING iii
HALAMAN PENGESAHAN PENGUJI iv
SURAT PERNYATAAN KARYA ASLI TUGAS AKHIR v
PERNYATAAN BEBAS PLAGIASI vi
ABSTRAK vii
ABSTRACT viii
KATA PENGANTAR ix
DAFTAR ISI x
DAFTAR TABEL xiii
DAFTAR GAMBAR xiv
DAFTAR LAMPIRAN xv
BAB I PENDAHULUAN 1
1.1 Latar Belakang Masalah 1
1.2 Rumusan Masalah 2
1.3 Batasan Masalah 2
1.4 Tujuan Penelitian 2
1.5 Manfaat Penelitian 3
1.6 Tahapan Penelitian 3
1.7 Sistematika Penulisan 4
BAB II TINJAUAN LITERATUR 5
2.1 Kanker Paru-paru 5
2.2 *Protein-Protein Interaction* 5
2.3 Pangkalan Data STRING 7
2.4 *Link Prediction* 7
2.4.1 Common Neighbors (CN) 8
2.4.2 Jaccard Coefficient (JC) 9
2.4.3 Adamic-Adar Index (AA) 9
2.4.4 *Resource Allocation Index* (RA) 9
2.4.5 *Preferential Attachment* (PA) 10
2.4.6 *Local Path Index* (*LP*) 10
2.4.7 L3 *Index* 10
2.5 Random Forest 11
2.6 Evaluasi 12
2.6.1 AUC-ROC 12
2.6.2 AUPR 13
2.6.3 *Accuracy* 13
2.6.4 *Top-k Precision* 13

2.7 Penelitian Terdahulu 14
BAB III METODOLOGI PENELITIAN 17
3.1 *Business Understanding* 18
3.2 Data *Understanding* 18
3.2.1 Sumber Data 19
3.2.2 Eksplorasi Data 20
3.3 Data *Preparation* 21
3.3.1 *Load* dan *Cleaning* Data 22
3.3.2 Filter *Confidence Score* 23
3.3.3 Pembangunan Graf PPI 24
3.3.4 *Largest Connected Component* (LCC) 24
3.3.5 *Split Training/Testing* dengan *Bridge Protection* 25
3.3.6 Negative Sampling 25
3.3.7 Pembuatan Data Latih *Random Forest* 26
3.4 Modeling 27
3.4.1 Metode Berbasis Indeks Kemiripan Topologis 28
3.4.2 *Random Forest* 32
3.5 Evaluasi 34
3.5.1 Metrik Evaluasi 34
3.6 *Deployment* 38
BAB IV HASIL PENGUJIAN DAN PEMBAHASAN 40
4.1 Implementasi 40
4.1.1 Deskripsi Alat dan Lingkungan Eksekusi 40
4.1.2 Tahapan Implementasi 41
4.2 Hasil 44
4.2.1 *Data Preparation* 44
4.2.2 Evaluasi Skenario *Medium Confidence* (threshold ≥ 0,40) 44
4.2.3 Evaluasi Skenario *High Confidence* (threshold ≥ 0,70) 45
4.2.4 Evaluasi Skenario *Highest Confidence* (threshold ≥ 0,90) 46
4.2.5 Perbandingan Antar Skenario 46
4.2.6 *Feature Importance Random Forest* 47
4.2.7 *Deployment* — Kandidat Interaksi Protein Baru 49
4.3 Pembahasan 50
4.3.1 Analisis Performa Komparatif Antar Metode 50
4.3.2 Pengaruh *Confidence Score* Terhadap Performa Metode 54
4.3.3 Analisis *Feature Importance* dan Relevansi L3 *Index* 54
4.3.4 Batasan Penelitian 55

BAB V PENUTUP 56
5.1 Kesimpulan 56
5.2 Saran 57
DAFTAR PUSTAKA 58
LAMPIRAN 60