DAFTAR ISI

ABSTRAK iii

ABSTRACT iv

LEMBAR PENGESAHAN v

LEMBAR PERNYATAAN ORISINALITAS vi

KATA PENGANTAR vii

DAFTAR ISI viii

DAFTAR GAMBAR xi

DAFTAR TABEL xii

DAFTAR LAMPIRAN xiii

DAFTAR SIMBOL xiv

DAFTAR ISTILAH xv

Bab I PENDAHULUAN 1

I.1 Latar Belakang 1

I.2 Perumusan Masalah 3

I.3 Tujuan Penelitian 3

I.4 Batasan Penelitian 4

I.5 Manfaat Penelitian 4

I.6 Metodologi Penelitian 4

Bab II TINJAUAN PUSTAKA 6

II.1 Penelitian Terdahulu 6

II.2 Dasar Teori 11

II.2.1 Penyakit Umum 11

II.2.2 Data Mining 11

II.2.3 CRISP-DM 12

II.2.4 Klasifikasi 12
II.2.5 Decision Tree 13
II.2.6 C4.5 14
II.2.7 CART 15
II.2.8 Random Over Under Sampling 16
II.2.9 Feature Selection 16
II.2.10 Hyperparameter 17
II.2.11 Sequential Searching 18
II.2.12 Confusion Matrix 18
II.2.13 Cross Validation 20
II.2.14 Sistem Diagnosis 20
Bab III METODOLOGI PENELITIAN 22
III.1 Metode yang Digunakan 22
III.1.1 Business Understanding 23
III.1.2 Data Understanding 23
III.1.3 Data Preparation 24
III.1.4 Modeling 25
III.1.5 Evaluation and Result 26
III.1.6 Deployment 29
III.2 Alat dan Bahan Penelitian 30
III.2.1 Google Colab 30
III.2.2 Visual Studio Code 30
III.2.3 Python 30
III.3 Prosedur Penelitian 31
III.3.1 Identifikasi Masalah 32
III.3.2 Studi Literatur 32

III.3.3 CRISP-DM 32
III.3.4 Penyusunan Buku TA 32
III.4 Jadwal Pelaksanaan 33
Bab IV ANALISIS DAN PERANCANGAN 34
IV.1 Business Understanding 34
IV.2 Data Understanding 35
IV.3 Data Preparation 37
IV.3.1 Data Cleaning 37
IV.3.2 Normalisasi Bahasa 41
IV.3.3 Random Over Under Sampling 41
IV.3.4 Transformasi Data 42
IV.3.5 Dataset Akhir 44
IV.4 Modeling 44
IV.4.1 Mencari Split Data Terbaik 42
IV.4.2 Distribusi Data 42
IV.4.3 Feature Selection 44
IV.4.4 Hyperparameter 45
IV.4.5 Perbandingan Model C4.5 dan CART 47
IV.5 Evaluation and Result 52
IV.5.1 Cross Validation C4.5 53
IV.5.2 Cross Validation CART 54
IV.6 Deployment 55
Bab V KESIMPULAN DAN SARAN 59
V.1 Kesimpulan 59
V.2 Saran 60
DAFTAR PUSTAKA 62