# DAFTAR ISI

**LEMBAR PENGESAHAN TUGAS AKHIR** i

**PERNYATAAN KEASLIAN DAN PERSETUJUAN PUBLIKASI TA** iii

**KATA PENGANTAR** v

**ABSTRAK** vii

**DAFTAR ISI** xi

**DAFTAR TABEL** xv

**DAFTAR GAMBAR** xvii

**DAFTAR LAMPIRAN** xxi

**BAB 1. PENDAHULUAN** 1

 1.1 Latar Belakang 1

1.2 Rumusan Masalah 2

1.3 Batasan Masalah 2

1.4 Tujuan Penelitian 2

**BAB 2. TINJAUAN PUSTAKA** 3

2.1. Diabetes 3

2.2. Machine Learning 4

2.3. Decision Tree 5

2.4. Gradient Boosting 6

2.5. K-NN 8

2.6. Logistic Regersion 9

2.7. Random Forest 10

2.8. Python 12

[2.9. Scikit-Learn](#_Toc27829444) 12

[2.10. Penelitian Terdahulu](#_Toc27829445) 13

**BAB 3.** [**METODE PENELITIAN**](#_Toc27829447) 15

[3.1. Perancangan Sistem](#_Toc27829448) 15

[3.1.1 Cara Kerja Machine Learning](#_Toc27829449) 15

[3.1.2 Kategori algoritma dan Teknik machine learning](#_Toc27829450) 16

[3.1.3 Cara Kerja Supervised Learning](#_Toc27829451) 17

[3.1.4 Flowchart Supervised Learning](#_Toc27829452) 18

[3.1.5 Flowchart Machine Learning Perbandingan Tingkat Akurasi Prediksi Diabetes](#_Toc27829453) 19

[3.1.6 Flowchart Metode K-NN](#_Toc27829454) 20

[3.1.7 Flowchart Metode Logistic Linear Regresion](#_Toc27829455) 21

[3.1.8 Flowchart Metode Decision Tree](#_Toc27829456) 21

[3.1.9 Flowchart Metode Random Forest](#_Toc27829457) 22

[3.1.10 Flowchart Metode Gradient Boosting](#_Toc27829458) 23

[3.2. Implementasi](#_Toc27829459) 23

[3.2.1 Tools](#_Toc27829460) 24

[3.2.2 Lingkungan Uji coba](#_Toc27829461) 24

**BAB 4.** [**HASIL DAN PEMBAHASAN**](#_Toc27829463) 27

[4.1. Uji Coba Sistem](#_Toc27829464) 27

[4.1.1 Alur proses dokumentasi import dataset dan overview data](#_Toc27829465) 27

[4.1.2 Hasil uji coba dengan metode K-NN](#_Toc27829466) 33

[4.1.3 Hasil uji coba dengan metode dengan Logistic-Regresion](#_Toc27829467) 34

[4.1.4 Hasil uji coba dengan metode dengan Decision-Tree](#_Toc27829468) 36

[4.1.5 Hasil uji coba dengan metode dengan Random Forest](#_Toc27829469) 38

[4.1.6 Hasil uji coba dengan metode dengan Gradient Boosting](#_Toc27829470) 40

[4.2. Dataset Yang Digunakan Dalam Pengujian](#_Toc27829471) 42

[4.3. Hasil Akurasi Tiap Metode](#_Toc27829472) 43

[4.4. Cross Validation dan Perhitungan Akurasi dan Missclasification Tiap Metode .](#_Toc27829473) 44

[4.4.1 Import library dan read data](#_Toc27829474) 44

[4.4.2 Spliting dataset](#_Toc27829475) 45

[4.4.3 Scaling feature](#_Toc27829476) 47

[4.4.4 Parameter Evaluation](#_Toc27829477) 48

[4.4.5 Adjusting Development Thereshold](#_Toc27829478) 51

[4.4.6 Predicting Tes set result or Storing prediction](#_Toc27829479) 54

[4.4.7 Making confusion Matrix](#_Toc27829480) 56

[4.4.8 Ploting prediction and result](#_Toc27829481) 60

[4.5. Hasil Dari Uji Coba](#_Toc27829482) 75

[4.5.1 Result uji coba akurasi](#_Toc27829483) 75

[4.5.2 Result uji coba dengan applikasi](#_Toc27829484) 76

**BAB 5.** [**PENUTUP**](#_Toc27829488) 85

[5.1. Kesimpulan](#_Toc27829489) 85

[5.2. Saran](#_Toc27829490) 85

[**DAFTAR PUSTAKA**](#_Toc27829491) 87

[**DAFTAR LAMPIRAN**](#_Toc27829492) 89

*Halaman ini sengaja dikosongkan*

# DAFTAR TABEL

[Tabel 3.1 Persiapan tools yang di gunakan](#_Toc27213341) 24

[Tabel 4.2 contoh dataset pima-indian-diabetes](#_Toc27213342) 29

[Tabel 4.3 dataset yang digunakan dalam pengujian akurasi prediksi supervised learning](#_Toc27213343) 45

[Tabel 4.4 Hasil akurasi prediksi diabetes tiap metode](#_Toc27213344) 46

[Tabel 5.5 hasil cross validation masing-masing metode](#_Toc27213345) 77

#

*Halaman ini sengaja dikosongkan*

# DAFTAR GAMBAR

[Gambar 2. 1 rumus gain entropy decision tree](file:///D%3A%5CTA%5CSidang%20akhir%5CLaporan%5Cfinal%5CLAPORAN%20SIDANG%20AKHIR%20B5%20-%20update.docx#_Toc27213363) 6

[Gambar 2.2 rumus mencari nilai entropy metode decision tree](#_Toc27213364) 6

[Gambar 2.3 rumus optimasi pemecahan metode Gradient Boosting](#_Toc27213365) 7

[Gambar 2.4 rumus peningkatan gradient metode Gradient Boosting](#_Toc27213366) 7

[Gambar 2.5 gambar rumus metode K-NN](#_Toc27213367) 9

[Gambar 2.6 gambar rumus regersi logistik](#_Toc27213368) 10

[Gambar 2.7 fungsi distribusi random vector](#_Toc27213369) 10

[Gambar 2.8 fungsi mencari eror metode random forest](#_Toc27213370) 11

[Gambar 2. 9 penggabungan 2 rumus eror dan distribusi metode random forest](#_Toc27213371) 11

[Gambar 2. 10 rumus mencari Kekuatan dan Korelasi metode Decision tree](#_Toc27213372) 11

[Gambar 3.1 cara kerja machine learning](file:///D%3A%5CTA%5CSidang%20akhir%5CLaporan%5Cfinal%5CLAPORAN%20SIDANG%20AKHIR%20B5%20-%20update.docx#_Toc27213373) 15

[Gambar 3.2 metode pemisahan data supervised dan unsupervised](file:///D%3A%5CTA%5CSidang%20akhir%5CLaporan%5Cfinal%5CLAPORAN%20SIDANG%20AKHIR%20B5%20-%20update.docx#_Toc27213374) 16

[Gambar 3**.**3 cara kerja supervised learning](file:///D%3A%5CTA%5CSidang%20akhir%5CLaporan%5Cfinal%5CLAPORAN%20SIDANG%20AKHIR%20B5%20-%20update.docx#_Toc27213375) 17

[Gambar 3.4 flowchart supervised learning](file:///D%3A%5CTA%5CSidang%20akhir%5CLaporan%5Cfinal%5CLAPORAN%20SIDANG%20AKHIR%20B5%20-%20update.docx#_Toc27213376) 18

[Gambar 3.5 flowchart alur sistem](file:///D%3A%5CTA%5CSidang%20akhir%5CLaporan%5Cfinal%5CLAPORAN%20SIDANG%20AKHIR%20B5%20-%20update.docx#_Toc27213377) 19

[Gambar 3.6 contoh flowcahrt metode K-NN](#_Toc27213378) 20

[Gambar 3.7 contoh Flowchart general data metode Linear Regeresion](#_Toc27213379) 21

[Gambar 3.8 flowchart metode decision tree dalam program python](#_Toc27213380) 21

[Gambar 3.9 flow metode Random Forest Using Python](#_Toc27213381) 22

[Gambar 3.10 flow metode Gradient Boosting](#_Toc27213382) 23

[Gambar 4.1 gambar source untuk load libray](#_Toc27213383) 28

[Gambar 4.2 gambar source untuk load dataset](#_Toc27213384) 28

[Gambar 4.3 gambar source untuk print data diabetes pima indian](#_Toc27213385) 29

[Gambar 4.4 gambar source untuk menampilakan dimensi data dan grup by data outcome](#_Toc27213386) 29

[Gambar 4.5 gambar untuk menjelaskan tentang basic EDA dari data outcome atau fitur label pada dataset](#_Toc27213387) 30

[Gambar 4.6 mengetahui info type data yang akan di uji](#_Toc27213388) 31

[Gambar 4.7 image pairplot data fitur](#_Toc27213389) 31

[Gambar 4.8 analisis data menggunakan diagram](#_Toc27213390) 32

[Gambar 4.9 spliting dataset uji coba pertama dengan random state = 66](#_Toc27213391) 33

[Gambar 4.10 hasil output visualization menggunakan metode K-NN](#_Toc27213392) 34

[Gambar 4.11 hasil akurasi prediksi dengan metode K-NN](#_Toc27213393) 34

[Gambar 4.12 hasil prediksi diabetes menggunakan metode Logistic-Regresion](#_Toc27213394) 35

[Gambar 4.13 hasil output visulization metode Logistic-Regresion](#_Toc27213395) 36

[Gambar 4.14 hasil prediksi diabetes menggunakan metode Decision-Tree](#_Toc27213396) 37

[Gambar 4.15 hasil output visualization metode Decision-Tree](#_Toc27213397) 38

[Gambar 4.16 hasil prediksi diabetes dengan metode Random Forest](#_Toc27213398) 39

[Gambar 4.17 hasil output visualization metode Random Forest](#_Toc27213399) 39

[Gambar 4.18 hasil prediksi diabetes menggunakan metode Gradient Boosting](#_Toc27213400) 40

[Gambar 4.19 hasil output visualization metode Gradient Boosting](#_Toc27213401) 41

[Gambar 4.20 import libaray dan read dataset metode Gradient boosting](#_Toc27213402) 45

[Gambar 4.21 perintah split dataset](#_Toc27213403) 46

[Gambar 4.22 hasil dari split dataset](#_Toc27213404) 46

[Gambar 4.23 Proses Scaling Feature](#_Toc27213405) 47

[Gambar 4.24 Proses Evaluasi Parameter metode Gradient Boosting](#_Toc27213406) 48

[Gambar 4.25 Proses Evaluasi Parameter metode Decision Tree](#_Toc27213407) 49

[Gambar 4.26 Proses Evaluasi Parameter metode K-NN](#_Toc27213408) 49

[Gambar 4.27 Proses Evaluasi Parameter metode Logistic Regresion](#_Toc27213409) 50

[Gambar 4.28 Proses Evaluasi Parameter metode Random Forest](#_Toc27213410) 50

[Gambar 4.29 Proses menetukan adusting development thereshold metode Gradient Boosting](#_Toc27213411) 51

[Gambar 4.30 Proses menetukan adusting development thereshold metode Decision Tree](#_Toc27213412) 52

[Gambar 4.31 Proses menetukan adusting development thereshold metode K-NN](#_Toc27213413) 52

[Gambar 4.32 Proses menetukan adusting development thereshold metode Logistic Regresion](#_Toc27213414) 53

[Gambar 4.33 Proses menetukan adusting development thereshold metode Random Forest](#_Toc27213415) 53

[Gambar 4.34 Proses prediction result metode Gradient Boosting](#_Toc27213416) 54

[Gambar 4.35 Proses prediction result metode Decision Tree](#_Toc27213417) 55

[Gambar 4.36 Proses prediction result metode K-NN](#_Toc27213418) 55

[Gambar 4.37 Proses prediction result metode Logistic Regresion](#_Toc27213419) 55

[Gambar 4.38 Proses prediction result metode Random Forest](#_Toc27213420) 56

[Gambar 4.39 hasil dari confusion matrix metode Gradient Boosting](#_Toc27213421) 57

[Gambar 4.40 hasil dari confusion matrik metode Decision Tree](#_Toc27213422) 58

[Gambar 4.41 hasil dari confusion matrik metode K-NN](#_Toc27213423) 58

[Gambar 4.42 hasil dari confusion matrik metode Logistic Regresion](#_Toc27213424) 59

[Gambar 4.43 hasil dari confusion matrik metode Random Forest](#_Toc27213425) 60

[Gambar 4.44 hasil akurasi prediksi metode Gradient Boosting](#_Toc27213426) 61

[Gambar 4.45 hasil akurasi prediksi metode Decision Tree](#_Toc27213427) 62

[Gambar 4.46 hasil akurasi prediksi metode K-NN](#_Toc27213428) 63

[Gambar 4.47 hasil akurasi prediksi metode Logistic Regresion](#_Toc27213429) 64

[Gambar 4.48 hasil akurasi prediksi metode Random Forest](#_Toc27213430) 65

[Gambar 4.49 hasil clasification report metode Gradient Boosting](#_Toc27213431) 66

[Gambar 4.50 hasil clasification report metode Decision Tree](#_Toc27213432) 67

[Gambar 4.51 hasil clasification report metode K-NN](#_Toc27213433) 68

[Gambar 4.52 hasil clasification report metode Logistic Regresion](#_Toc27213434) 69

[Gambar 4.53 hasil clasification report metode Random Forest](#_Toc27213435) 70

[Gambar 4.54 grafik result roc auc metode Gradient Boosting](#_Toc27213436) 72

[Gambar 4.55 grafik result roc auc metode Decision Tree](#_Toc27213437) 72

[Gambar 4.56 grafik result roc auc metode K-NN](#_Toc27213438) 73

[Gambar 4.57 grafik result roc auc metode Logistic Regresion](#_Toc27213439) 74

[Gambar 4.58 grafik result roc auc metode Random Forest](#_Toc27213440) 74

[Gambar 4.59 grafik kesimpulan perbandingan akurasi 5 metode](#_Toc27213441) 76

[Gambar 4.60 running front end framework Laravel port 8000](#_Toc27213442) 77

[Gambar 4.61 running back end python-fask port 5000](#_Toc27213443) 77

[Gambar 4.62 tampilan awal menu](#_Toc27213444) 78

[Gambar 4.63 menu input data](#_Toc27213445) 79

[Gambar 4.64 contoh input data](#_Toc27213446) 80

[Gambar 4.65 result atau output dari input data](#_Toc27213447) 81

[Gambar 4.66 result you don’t have a risk of diabetes](#_Toc27213448) 81

*Halaman ini sengaja dikosongkan*

# DAFTAR LAMPIRAN

[Lampiran 1 : documentasi comparasion metode di google colab](#_Toc27213449) 93

[Lampiran 2 : dokumentasi pengembangan akurasi metode dengan deep learning](#_Toc27213450) 94

[Lampiran 3 : dokumentasi deploy machine learning di kaagle](#_Toc27213451) 95

[Lampiran 4 : pengembangan program machine learnig apps](#_Toc27213452) 97

[Lampiran 5 publikasi jurnal lolos seleksi kompetisi BICMST “Brawijaya International Conference Multidisiplionary Sience and technology”](#_Toc27213453) 98

*Halaman ini sengaja dikosongkan*