

DAFTAR ISI

HALAMAN Sampul	i
HALAMAN Judul.....	ii
HALAMAN Persetujuan.....	iii
HALAMAN Pengesahan.....	iv
HALAMAN Pernyataan Keaslian	v
HALAMAN Persembahan	vi
HALAMAN Motto	vii
Kata Pengantar	viii
DAFTAR ISI.....	ix
DAFTAR Tabel.....	xi
DAFTAR Gambar	xiv
DAFTAR Istilah	xix
DAFTAR Lampiran.....	xx
Intisari.....	xxi
ABSTRACT.....	xxii
BAB I PENDAHULUAN	
A. Latar Belakang Masalah	1
B. Rumusan Masalah.....	6
C. Batasan Masalah	6
D. Tujuan Penelitian	7
E. Manfaat Penelitian	7
BAB II TINJAUAN PUSTAKA	
A. Landasan Teori.....	8
1. <i>Machine Learning</i>	8
2. <i>Confusion Matrix</i>	14
3. Kementerian ATR/BPN	17
4. PTSL.....	20
5. <i>Phyton</i>	22
6. <i>Google Colaboratory</i>	23
B. Penelitian Sebelumnya.....	24

BAB III METODE PENELITIAN

A. Tempat dan Waktu Penelitian.....	28
B. Metode Pengumpulan Data.....	28
1. Observasi.....	28
2. Studi Pustaka.....	28
C. Alat dan Bahan Penelitian.....	29
1. Alat.....	23
2. Bahan.....	30
D. Konsep Penelitian.....	30
1. <i>Data Collecting</i>	31
2. <i>Preprocessing Data</i>	31
3. Membangun Model <i>Machine Learning</i>	33
4. Implementasi.....	40
5. Evaluasi.....	41

BAB IV HASIL DAN PEMBAHASAN

A. <i>Data Collecting</i>	45
B. <i>Preprocessing Data</i>	46
1. Persiapan Data.....	46
2. Penanganan <i>Missing Value</i>	51
3. Seleksi Fitur.....	53
4. Pembagian <i>Training Set</i> dan <i>Testing Set</i>	57
C. Membangun Model <i>Machine Learning</i>	60
D. Implementasi.....	66
E. Evaluasi.....	68
1. Evaluasi Pengujian Skenario <i>Splitting Data</i>	68
2. Evaluasi menggunakan studi kasus sederhana.....	151

BAB V PENUTUP

A. Kesimpulan.....	155
B. Saran.....	156

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

Tabel 2.1. Tabel Penelitian Sebelumnya.....	8
Tabel 4.1. Data Partisipasi masyarakat yang mendaftar dan tidak mendaftar dalam pembuatan sertifikat pada program PTSL Tahun 2023.....	45
Tabel 4.2. Dataset yang digunakan sesuai seleksi fitur.....	54
Tabel 4.3. Hasil Implementasi Label <i>Encoder</i> pada kolom Pekerjaan	55
Tabel 4.4. Hasil Implementasi Label <i>Encoder</i> pada kolom Penggunaan Tanah ...	56
Tabel 4.5. Hasil Implementasi Label <i>Encoder</i> pada kolom Keterangan.....	56
Tabel 4.6. Hasil <i>confusion matrix KNN</i> Pengujian 0.1 (<i>training</i> 90% dan jumlah data 10% untuk data uji)	70
Tabel 4.7. Hasil <i>confusion matrix SVM</i> pengujian 0.1 (<i>training</i> 90% dan jumlah data 10% untuk data uji)	72
Tabel 4.8. Hasil <i>confusion matrix Random Forest</i> pengujian 0.1 (<i>training</i> 90% dan jumlah data 10% untuk data uji)	73
Tabel 4.9. Hasil <i>confusion matrix decision tree (C5.0)</i> pengujian 0.1 (<i>training</i> 90% dan jumlah data 10% untuk data uji).....	75
Tabel 4.10. Hasil <i>confusion matrix KNN</i> pengujian 0.2 (<i>training</i> 80% dan jumlah data 20% untuk data uji) algoritma <i>KNN</i>	77
Tabel 4.11. Hasil <i>confusion matrix SVM</i> pengujian 0.2 (<i>training</i> 80% dan jumlah data 20% untuk data uji).....	80
Tabel 4.12. Hasil <i>confusion matrix Random Forest</i> (<i>training</i> 80% dan jumlah data 20% untuk data uji)	81
Tabel 4.13. Hasil <i>confusion matrix decision tree (C5.0)</i> pengujian 0.2 (<i>training</i> 80% dan jumlah data 20% untuk data uji).....	83
Tabel 4.14. Hasil <i>confusion matrix KNN</i> Pengujian 0.3 (<i>training</i> 70% dan jumlah data 30% untuk data uji).....	85
Tabel 4.15. Hasil <i>confusion matrix SVM</i> pengujian 0.3 (<i>training</i> 70% dan jumlah data 30% untuk data uji).....	87
Tabel 4.16. Hasil <i>confusion matrix random forest</i> pengujian 0.3 (<i>training</i> 70% dan jumlah data 30% untuk data uji).....	89
Tabel 4.17. Hasil <i>confusion matrix decision tree (C5.0)</i> pengujian 0.3 (<i>training</i> 70% dan jumlah data 30% untuk data uji).....	91
Tabel 4.18. Hasil <i>confusion matrix KNN</i> pengujian 0.4 (<i>training</i> 60% dan jumlah data 40% untuk data uji)	93
Tabel 4.19. Hasil <i>confusion matrix SVM</i> pengujian 0.4 (<i>training</i> 60% dan jumlah data 40% untuk data uji)	95
Tabel 4.20. Hasil <i>confusion matrix Random Forest</i> pengujian 0.4 (<i>training</i> 60% dan jumlah data 40% untuk data uji).....	97

Tabel 4.21. Hasil <i>confusion matrix decision tree</i> (C5.0) pengujian 0.4 (<i>training</i> 60% dan jumlah data 40% untuk data uji).....	99
Tabel 4.22. Hasil <i>confusion matrix KNN</i> Pengujian 0.5 (<i>training</i> 50% dan jumlah data 50% untuk data uji)	102
Tabel 4.23. Hasil <i>confusion matrix SVM</i> pengujian 0.5 (<i>training</i> 50% dan jumlah data 50% untuk data uji)	103
Tabel 4. 24. Hasil <i>confusion matrix random forest</i> dengan pengujian 0.5 (<i>training</i> 50% dan jumlah data 50% untuk data uji).....	105
Tabel 4.25. Hasil <i>confusion matrix decision tree</i> (C5.0) pengujian 0.5 (<i>training</i> 50% dan jumlah data 50% untuk data uji).....	107
Tabel 4.26. Hasil <i>confusion matrix KNN</i> pengujian 0.6 (<i>training</i> 60% dan jumlah data 40% untuk data uji)	110
Tabel 4.27. Hasil <i>confusion matrix SVM</i> pengujian 0.6 (<i>training</i> 60% dan jumlah data 40% untuk data uji)	112
Tabel 4.28. Hasil <i>confusion matrix random fssorest</i> pengujian 0.6 (<i>training</i> 60% dan jumlah data 40% untuk data uji).....	113
Tabel 4.29. Hasil <i>confusion matrix decision tree</i> (C5.0) pengujian 0.6 (<i>training</i> 60% dan jumlah data 40% untuk data uji)	115
Tabel 4.30. Hasil <i>confusion matrix KNN</i> pengujian 0.7 (<i>training</i> 70% dan jumlah data 30% untuk data uji)	118
Tabel 4.31. Hasil <i>confusion matrix SVM</i> pengujian 0.7 (<i>training</i> 70% dan jumlah data 30% untuk data uji)	120
Tabel 4.32. Hasil <i>confusion matrix Random Forest</i> pengujian 0.7 (<i>training</i> 70% dan jumlah data 30% untuk data uji).....	121
Tabel 4.33. Hasil <i>confusion matrix decision tree</i> (C5.0) pengujian 0.7 (<i>training</i> 70% dan jumlah data 30% untuk data uji).....	123
Tabel 4.34. Hasil <i>confusion matrix KNN</i> pengujian 0.8 (<i>training</i> 80% dan jumlah data 20% untuk data uji)	126
Tabel 4.35. Hasil <i>confusion matrix SVM</i> pengujian 0.8 (<i>training</i> 80% dan jumlah data 20% untuk data uji)	128
Tabel 4.36. Hasil <i>confusion matrix random forest</i> pengujian 0.8 (<i>training</i> 80% dan jumlah data 20% untuk data uji).....	129
Tabel 4.37. Hasil <i>confusion matrix decision tree</i> (C5.0) pengujian 0.8 (<i>training</i> 80% dan jumlah data 20% untuk data uji).....	131
Tabel 4.38. Hasil <i>confusion matrix KNN</i> pengujian 0.9 (<i>training</i> 90% dan jumlah data 10% untuk data uji)	134
Tabel 4.39. Hasil <i>confusion matrix SVM</i> pengujian 0.9 (<i>training</i> 90% dan jumlah data 10% untuk data uji)	136
Tabel 4. 40. Hasil <i>confusion matrix Random Forest</i> pengujian 0.1 (<i>training</i> 90% dan jumlah data 10% untuk data uji).....	137

Tabel 4.41. Hasil <i>confusion matrix decision tree</i> (C5.0) pengujian 0.1 (<i>training</i> 90% dan jumlah data 10% untuk data uji).....	139
Tabel 4.42. Tabel perbandingan skor prediksi pengujian 0.1 s/d 0.9 metode <i>KNN</i> , <i>SVM</i> , <i>random forest</i> dan <i>decision tree</i> (C5.0).....	141
Tabel 4.43. Perbandingan kecepatan program pada metode <i>SVM</i> , <i>random forest</i> dan <i>decision tree</i> (C5.0)	150
Tabel 4.44. Kesimpulan perbandingan kecepatan program pada metode <i>SVM</i> , <i>random forest</i> dan <i>decision tree</i> (C5.0).....	151
Tabel 4.45. Nilai Indeks partisipasi masyarakat dengan variabel yang sudah dirubah menjadi angka.....	151
Tabel 4.46. Tabel studi kasus partisipasi masyarakat dengan variabel yang sudah dirubah menjadi angka.....	152
Tabel 4.47. Data asli partisipasi masyarakat.....	153



DAFTAR GAMBAR

Gambar 2.1. Ilustrasi <i>hyperplane SVM</i>	11
Gambar 2.2. Ilustrasi <i>confusion matrix</i>	15
Gambar 3.1. Aplikasi <i>Excel</i>	29
Gambar 3.2. <i>Google Colab</i>	29
Gambar 3.3. Diagram Alur Penelitian.....	31
Gambar 3. 4. <i>Confusion Matrix</i>	41
Gambar 4.1. Mengambil dataset yang tersimpan pada <i>google drive</i>	47
Gambar 4.2. Dataset pada <i>google drive</i> sudah tampil di <i>google colab</i>	47
Gambar 4.3. <i>Import library</i> pada <i>google colab</i>	48
Gambar 4.4. Memanggil dataset pada <i>google colab</i>	48
Gambar 4.5. Tampilan data <i>frame</i> yang dipanggil.....	49
Gambar 4.6. Menampilkan info yang terdapat dalam <i>dataset</i>	50
Gambar 4.7. Jenis data <i>missing value</i>	51
Gambar 4.8. Jumlah data <i>missing value</i>	51
Gambar 4.9. Mengisi <i>missing value</i> dengan data <i>mean</i>	52
Gambar 4.10. Implementasi <i>Label Encoder</i> pada kolom Pekerjaan	55
Gambar 4.11. Implementasi <i>Label Encoder</i> pada kolom Penggunaan Tanah	56
Gambar 4.12. Implementasi <i>Label Encoder</i> pada kolom Keterangan	56
Gambar 4.13. Perubahan nilai variabel Pekerjaan, Penggunaan Tanah, Keterangan.....	57
Gambar 4.14. Menentukan variable x (variable independent).....	58
Gambar 4.15. Menentukan variable y (variable dependen)	58
Gambar 4.16. Melihat jumlah label “0” dan label “1”	59
Gambar 4.17. Membagi <i>train</i> dan <i>test set</i> dari <i>dataframe</i>	59
Gambar 4.18. Algoritma <i>SVM</i>	60
Gambar 4.19. Tampilan nilai <i>_y_pred_svms</i>	61
Gambar 4.20. Membangun metode <i>KNN</i>	61
Gambar 4.21. Tampilan nilai <i>_y_pred_knns</i>	62
Gambar 4.22. Membangun model <i>random forest</i>	63

Gambar 4.23. Tampilan nilai <i>y_pred_rfs</i>	64
Gambar 4.24. Membangun model <i>decision tree</i> (C5.0).....	65
Gambar 4. 25. Tampilan nilai <i>y_pred_c5s</i>	65
Gambar 4.26. Hasil data variabel yang sudah dirubah menjadi angka	66
Gambar 4. 27. Implementasi menggunakan model <i>SVM</i>	66
Gambar 4.28. Implementasi menggunakan model <i>KNN</i>	67
Gambar 4.29. Implementasi menggunakan model <i>random forest</i>	67
Gambar 4.30. Implementasi menggunakan model <i>decesion tree</i> (C5.0)	67
Gambar 4.31. Pengujian 0.1 (<i>training</i> 90% dan jumlah data 10% untuk data uji)	69
Gambar 4.32. Pengujian 0.1 (<i>training</i> 90% dan jumlah data 10% untuk data uji) algoritma <i>KNN</i>	69
Gambar 4.33. Pengujian 0.1 (<i>training</i> 90% dan jumlah data 10% untuk data uji) algoritma <i>SVM</i>	71
Gambar 4.34. Pengujian 0.1 (<i>training</i> 90% dan jumlah data 10% untuk data uji) algoritma <i>random forest</i>	73
Gambar 4.35. Pengujian 0.1 (<i>training</i> 90% dan jumlah data 10% untuk data uji) algoritma <i>decision tree</i> (C5.0).....	75
Gambar 4.36. Pengujian 0.2 (<i>training</i> 80% dan jumlah data 20% untuk data uji)	77
Gambar 4.37. Pengujian 0.2 (<i>training</i> 80% dan jumlah data 20% untuk data uji) algoritma <i>KNN</i>	77
Gambar 4.38. Pengujian 0.2 (<i>training</i> 80% dan jumlah data 20% untuk data uji) algoritma <i>SVM</i>	79
Gambar 4.39. Pengujian 0.2 (<i>training</i> 80% dan jumlah data 20% untuk data uji) algoritma <i>random forest</i>	81
Gambar 4.40. Pengujian 0.2 (<i>training</i> 80% dan jumlah data 20% untuk data uji) algoritma <i>decision tree</i> (C5.0).....	83
Gambar 4.41. Pengujian 0.3 (<i>training</i> 70% dan jumlah data 30% untuk data uji)	84

Gambar 4.42. Pengujian 0.3 (<i>training</i> 70% dan jumlah data 30% untuk data uji) algoritma <i>KNN</i>	85
Gambar 4.43. Pengujian 0.3 (<i>training</i> 70% dan jumlah data 30% untuk data uji) algoritma <i>SVM</i>	87
Gambar 4.44. Pengujian 0.3 (<i>training</i> 70% dan jumlah data 30% untuk data uji) algoritma <i>random forest</i>	89
Gambar 4.45. Pengujian 0.3 (<i>training</i> 70% dan jumlah data 30% untuk data uji) algoritma <i>decision tree</i> (C5.0).....	91
Gambar 4.46. Pengujian 0.4 (<i>training</i> 60% dan jumlah data 40% untuk data uji)	92
Gambar 4.47. Pengujian 0.4 (<i>training</i> 60% dan jumlah data 40% untuk data uji) algoritma <i>KNN</i>	93
Gambar 4.48. Pengujian 0.4 (<i>training</i> 60% dan jumlah data 40% untuk data uji) algoritma <i>SVM</i>	95
Gambar 4.49. Pengujian 0.4 (<i>training</i> 60% dan jumlah data 40% untuk data uji) algoritma <i>random forest</i>	97
Gambar 4.50. Pengujian 0.4 (<i>training</i> 60% dan jumlah data 40% untuk data uji) algoritma <i>decision tree</i> (C5.0).....	99
Gambar 4.51. Pengujian 0.5 (<i>training</i> 50% dan jumlah data 50% untuk data uji)	101
Gambar 4.52. Pengujian 0.5 (<i>training</i> 50% dan jumlah data 50% untuk data uji) algoritma <i>KNN</i>	101
Gambar 4.53. Pengujian 0.5 (<i>training</i> 50% dan jumlah data 50% untuk data uji) algoritma <i>SVM</i>	103
Gambar 4.54. Pengujian 0.5 (<i>training</i> 50% dan jumlah data 50% untuk data uji) algoritma <i>random forest</i>	105
Gambar 4.55. Pengujian 0.5 (<i>training</i> 50% dan jumlah data 50% untuk data uji) algoritma <i>decision tree</i> (C5.0).....	107
Gambar 4.56. Pengujian 0.6 (<i>training</i> 60% dan jumlah data 40% untuk data uji)	109

Gambar 4.57. Pengujian 0.6 (<i>training</i> 60% dan jumlah data 40% untuk data uji) algoritma <i>KNN</i>	109
Gambar 4.58. Pengujian 0.6 (<i>training</i> 60% dan jumlah data 40% untuk data uji) algoritma <i>SVM</i>	111
Gambar 4.59. Pengujian algoritma 0.6 (<i>training</i> 60% dan jumlah data 40% untuk data uji) <i>random forest</i>	113
Gambar 4.60. Pengujian 0.6 (<i>training</i> 60% dan jumlah data 40% untuk data uji) algoritma <i>decision tree</i> (C5.0).....	115
Gambar 4.61. Pengujian 0.7 (<i>training</i> 70% dan jumlah data 30% untuk data uji)	117
Gambar 4.62. Pengujian 0.7 (<i>training</i> 70% dan jumlah data 30% untuk data uji) algoritma <i>KNN</i>	117
Gambar 4.63. Pengujian 0.7 (<i>training</i> 70% dan jumlah data 30% untuk data uji) algoritma <i>SVM</i>	119
Gambar 4.64. Pengujian 0.7 (<i>training</i> 70% dan jumlah data 30% untuk data uji) algoritma <i>random forest</i>	121
Gambar 4.65. Pengujian algoritma 0.7 (<i>training</i> 70% dan jumlah data 30% untuk data uji) <i>decision tree</i> (C5.0)	123
Gambar 4.66. Pengujian 0.8 (<i>training</i> 80% dan jumlah data 20% untuk data uji)	125
Gambar 4.67. Pengujian 0.8 (<i>training</i> 80% dan jumlah data 20% untuk data uji) algoritma <i>KNN</i>	125
Gambar 4.68. Pengujian 0.8 (<i>training</i> 80% dan jumlah data 20% untuk data uji) algoritma <i>SVM</i>	127
Gambar 4.69. Pengujian 0.8 (<i>training</i> 80% dan jumlah data 20% untuk data uji) algoritma <i>random forest</i>	129
Gambar 4.70. Pengujian 0.8 (<i>training</i> 80% dan jumlah data 20% untuk data uji) algoritma <i>decision tree</i> (C5.0).....	131
Gambar 4.71. Pengujian 0.9 (<i>training</i> 90% dan jumlah data 10% untuk data uji)	133

Gambar 4.72. Pengujian 0.9 (<i>training</i> 90% dan jumlah data 10% untuk data uji) algoritma <i>KNN</i>	133
Gambar 4.73. Pengujian 0.9 (<i>training</i> 90% dan jumlah data 10% untuk data uji) algoritma <i>SVM</i>	135
Gambar 4.74. Pengujian 0.9 (<i>training</i> 90% dan jumlah data 10% untuk data uji) algoritma <i>random forest</i>	137
Gambar 4.75. Pengujian 0.9 (<i>training</i> 90% dan jumlah data 10% untuk data uji) algoritma <i>decision tree</i> (C5.0).....	139
Gambar 4.76. <i>Library</i> yang dibutuhkan dalam menghitung kecepatan komputasi.....	142
Gambar 4.77. Proses percobaan 1 menghitung kecepatan algoritma <i>SVM</i>	143
Gambar 4.78. Proses percobaan 1 menghitung kecepatan algoritma <i>random forest</i>	144
Gambar 4.79. Proses percobaan 1 menghitung kecepatan algoritma <i>decision tree</i> (C5.0)	145
Gambar 4.80. Proses percobaan 2 menghitung kecepatan algoritma <i>SVM</i>	146
Gambar 4.81. Proses percobaan 2 menghitung kecepatan algoritma <i>random forest</i>	146
Gambar 4.82. Proses percobaan 2 menghitung kecepatan algoritma <i>decision tree</i> (C5.0)	147
Gambar 4.83. Proses percobaan 3 menghitung kecepatan algoritma <i>SVM</i>	147
Gambar 4.84. Proses percobaan 3 menghitung kecepatan algoritma <i>random forest</i>	148
Gambar 4.85. Proses percobaan 3 menghitung kecepatan algoritma <i>decision tree</i> (C5.0)	148
Gambar 4.86. Proses percobaan 4 menghitung kecepatan algoritma <i>SVM</i>	149
Gambar 4.87. Proses percobaan 4 menghitung kecepatan algoritma <i>random forest</i>	149
Gambar 4.88. Proses percobaan 4 menghitung kecepatan algoritma <i>decision tree</i> (C5.0)	150