

**PERBANDINGAN KLASIFIKASI  
DATA PENEMPATAN PEKERJA MIGRAN INDONESIA  
PADA KANTOR BP3TKI SEMARANG  
ANTARA ALGORITMA C4.5 DAN ALGORITMA NAIVE BAYES**

Tugas Akhir disusun untuk memenuhi syarat  
mencapai gelar Kesarjanaan Komputer pada  
Program Studi Sistem Informasi  
Jenjang Program Strata – 1



**SAUFIKA SUKMAWATI**

**18.01.55.5004**

**20780**

**FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS STIKUBANK (UNISBANK)**

**SEMARANG**

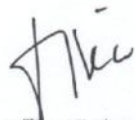
**2020**

## PERNYATAAN KESIAPAN UJIAN TUGAS AKHIR

Saya, Saufika Sukmawati, dengan ini menyatakan bahwa Laporan Tugas Akhir yang berjudul :

**Perbandingan Klasifikasi Data Penempatan Pekerja Migran Indonesia Pada Kantor BP3TKI Semarang Antara Algoritma C4.5 Dan Algoritma Naive Bayes**

Adalah benar hasil karya saya dan belum pernah diajukan sebagai karya ilmiah, sebagian atau seluruhnya, atas nama saya atau pihak lain.



(Saufika Sukmawati)  
NIM : 18.01.55.5004

Disetujui oleh Pembimbing :

Kami setuju Laporan tersebut diajukan untuk Ujian Tugas Akhir

Semarang : 17 Juni 2020



(Dra. SULASTRI, M.Kom)  
NIDN. 0608126601  
Pembimbing



# UNIVERSITAS STIKUBANK "UNISBANK" SEMARANG

## FAKULTAS TEKNOLOGI INFORMASI

Rektorat Kampus Mugas :  
Jl. Tri Lomba Juang No. 3 Semarang 50241  
Telp. (024) 8451976, 8311668, 8454746, Fax (024) 8443240  
E-mail : info@unisbank.ac.id

Kampus Kendeng :  
Jl. Kendeng V Bendan Ngisor Semarang  
Telp. (024) 8414970, Fax (024) 8441738  
E-mail : fe@unisbank.ac.id

### SURAT PERNYATAAN KEASLIAN TUGAS AKHIR/ SKRIPSI

Yang bertanda tangan di bawah ini, saya menyatakan bahwa TUGAS AKHIR / SKRIPSI dengan Judul :

**PERBANDINGAN KLASIFIKASI DATA PENEMPATAN PEKERJA MIGRAN INDONESIA PADA KANTOR BP3TKI SEMARANG ANTARA ALGORITMA C4.5 DAN ALGORITMA NAIVE BAYES**

yang telah diuji di depan tim penguji pada tanggal 28 Juli 2020, adalah benar hasil karya saya dan dalam TUGAS AKHIR /SKRIPSI ini tidak terdapat keseluruhan atau sebagian tulisan orang lain yang saya ambil dengan cara menyalin, atau meniru dalam bentuk rangkaian kalimat atau simbol yang saya aku seolah-olah sebagai tulisan saya sendiri dan atau tidak terdapat bagian atau keseluruhan tulisan yang saya salin, tiru atau yang saya ambil dari tulisan orang lain tanpa memberikan pengakuan pada penulis aslinya.

Apabila saya melakukan hal tersebut diatas, baik sengaja maupun tidak, dengan ini saya menyatakan menarik TUGAS AKHIR / SKRIPSI yang saya ajukan sebagai hasil tulisan saya sendiri.

Bila kemudian terbukti bahwa saya ternyata melakukan tindakan menyalin atau meniru tulisan orang lain seolah-olah hasil pemikiran saya sendiri, berarti gelar dan ijazah saya yang telah diberikan oleh Universitas Stikubank (UNISBANK) Semarang batal saya terima.

Semarang, 4 Agustus 2020

Yang Menyatakan



**( SAUFIKA SUKMAWATI )**

NIM : 18.01.55.5004

SAKSI 1

Tim Penguji

**( SULASTRI, DRA, M.KOM )**

SAKSI 2

Tim Penguji

**( ARIEF JANANTO, S.Kom., M.Cs. )**

SAKSI 3

Tim Penguji

**( SUGIYAMTA, S.Kom, M.Kom )**

## HALAMAN PENGESAHAN

Tugas Akhir dengan judul:

PERBANDINGAN KLASIFIKASI DATA PENEMPATAN  
PEKERJA MIGRAN INDONESIA PADA KANTOR BP3TKI SEMARANG  
ANTARA ALGORITMA C4.5 DAN ALGORITMA NAIVE BAYES

Ditulis oleh :

NIM : 18.01.55.5004

Nama : Saufika Sukmawati

Telah dipertahankan di depan Tim Dosen Penguji Tugas Akhir dan diterima sebagai salah satu syarat guna menyelesaikan Jenjang Program Strata 1 Program Studi Sistem Informasai pada Fakultas Teknologi Informasi Universitas Stikubank (UNISBANK) Semarang.

Semarang, 4 Agustus 2020

Ketua



(Dra. Sulastri, M.Kom.)

NIDN : 0608126601

Sekretaris



(Arief Jananto, S.Kom., M.Cs.)

NIDN : 0606017402

Anggota



(Sugiyamta, M.Kom.)

NIDN : 0615076702

Mengetahui,  
Universitas Stikubank (UNISBANK) Semarang  
Fakultas Teknologi Informasi  
Dekan



(Kristophorus Hadiono, S.Kom., Ph.D)

NIDN : 0622027601



**FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS STIKUBANK (UNISBANK) SEMARANG**

Program Studi : Sistem Informasi

Tugas Akhir Sarjana Komputer

Semester Genap tahun 2019

**PERBANDINGAN KLASIFIKASI DATA PENEMPATAN PEKERJA MIGRAN  
INDONESIA PADA KANTOR BP3TKI SEMARANG  
ANTARA ALGORITMA C4.5 DAN ALGORITMA NAIVE BAYES**

**Saufika Sukmawati**

**18.01.55.5004**

**Abstrak**

Penempatan tenaga kerja Indonesia ke luar negeri merupakan salah satu upaya pemerintah dalam mewujudkan hak masyarakat untuk mendapatkan kesempatan bekerja serta meningkatkan perekonomian negara. Sebagai salah satu upaya perlindungan pekerja migran Indonesia maka dikembangkan sebuah sistem komputerisasi tenaga kerja luar negeri (SISKOTKLN) oleh Badan Nasional Penempatan dan Perlindungan TKI (BNP2TKI).

Permasalahan yang menjadi kendala adalah adanya pekerja migran Indonesia (PMI) yang dipulangkan atau mendapat permasalahan ketenagakerjaan selama diluar negeri. Sehingga dibutuhkan sebuah interpretasi pada pola data penempatan PMI yang dapat digunakan sebagai prediksi negara tujuan penempatan para calon PMI kedepannya.

Data yang digunakan dalam penelitian adalah data penempatan PMI pada wilayah BP3TKI Semarang dengan skema penempatan antara agency Indonesia dan agensi di negara tujuan. Penulis membandingkan dua algoritma klasifikasi dalam data mining yaitu algoritma C 4.5 dan algoritma Naïve Bayes. Percobaan dengan data training sebanyak 1802 dan data testing sebanyak 772 menghasilkan nilai akurasi paling tinggi bagi kedua algoritma. Algoritma C 4.5 mampu memprediksi lebih baik dengan tingkat akurasi sebesar 84.84% sedangkan Algoritma Naive Bayes menghasilkan nilai akurasi sebesar 58.29%.

**Kata Kunci:** *C 4.5, Naïve Bayes, pekerja migran Indonesia*

**Abstract**

*The placement of Indonesian migrant workers abroad is one of the government's efforts to make people get more chance to get job opportunities as well as improving their livelihood. The National Board on The Placement and Protection of Indonesian Migrant Workers (BNP2TKI), as a government organization has*

*developed digital based system called SISKOTKLN as an aegis for Indonesian migrant workers.*

*The problem becomes an obstacle several Indonesian migrant workers have labor problems and some of them have been repatriated from their placement country. An interpretation of the migrant worker placement data pattern is needed which can be used as a prediction of the destination countries for migrant worker candidates ahead.*

*This research focused on Indonesian migrant worker placement data from the BP3TKI Semarang office with private to private placement schemes which are facilitated by Private companies in both Indonesia and destination countries. This research compares the C 4.5 algorithm and the Naïve Bayes classifier which are the two classification algorithms in data mining. The highest accuracy values for both algorithms produced by using 1802 training data and 772 testing data. The C 4.5 algorithm can predict better with an accuracy rate of 84.84% while the Naive Bayes algorithm produces an accuracy value of 58.29%.*

**Keywords:** *C 4.5, Naïve Bayes, Indonesian migrant worker*

Semarang, 17 Juli 2020  
Pembimbing

  
**Dra. SULASTRI, M.Kom**  
NIDN. 0608126601

## **KATA PENGANTAR**

Segala puji dan syukur atas kehadiran Allah SWT yang senantiasa diberikan kepada penulis, sehingga penulis dapat menyelesaikan tugas akhir dengan judul “PERBANDINGAN KLASIFIKASI DATA PENEMPATAN PEKERJA MIGRAN INDONESIA PADA KANTOR BP3TKI SEMARANG ANTARA ALGORITMA C 4.5 DAN ALGORITMA NAÏVE BAYES” sebagai syarat dalam menyelesaikan Jenjang Program Strata 1 Program Studi Teknik Informatika pada Fakultas Teknologi Informasi Universitas Stikubank Semarang.

Selama penulisan laporan dan penyelesaian tugas akhir ini penulis ingin mengucapkan rasa terima kasih yang sebesar-besarnya kepada:

1. Dr.Safik Faozi, SH, M.Hum selaku Rektor Universitas Stikubank Semarang.
2. Kristophorus Hadiono, S.Kom., Ph.D selaku Dekan Fakultas Teknologi Informasi Universitas Stikubank Semarang
3. Arief Jananto, S.Kom., M.Cs. selaku Ketua Program Studi Sistem Informasi Universitas Stikubank Semarang
4. Dra. Sulastri M.Kom selaku Dosen Pembimbing yang sudah berkenan meluangkan waktunya untuk memberikan bimbingan dan arahan selama penyusunan tugas akhir.
5. Seluruh Dosen dan Staf Fakultas Teknologi Informasi Universitas Stikubank Semarang.

6. Seluruh staf BP3TKI Semarang yang sudah memberikan izin penelitian dan membantu dalam kelancaran penelitian ini.
7. Kedua orang tua yang selalu memberikan doa dan dukungan penuh kepada penulis selama menempuh kuliah di program studi Sistem Informasi, dan
8. Rekan – rekan Sistem Informasi dan Fakultas Teknologi Informasi yang selalu menginspirasi penulis dalam menyelesaikan tugas akhir ini.

Akhir kata penulis berharap semoga tugas akhir ini dapat bermanfaat dan mendorong penelitian-penelitian berikutnya.

Semarang, Juni 2020  
Penulis

Saufika Sukmawati  
NIM.18.01.5.5004



## DAFTAR ISI

HALAMAN JUDUL .....	ii
HALAMAN KESIAPAN UJIAN TUGAS AKHIR .....	iii
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR.....	iv
HALAMAN PENGESAHAN.....	v
ABSTRAK .....	vi
KATA PENGANTAR .....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR .....	xiv
DAFTAR RUMUS.....	xvi
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang Penelitian .....	1
1.2 Perumusan Masalah .....	3
1.3 Tujuan dan Manfaat Penelitian .....	3
1.4 Metode Penelitian .....	4
1.5 Sistematika Penulisan .....	5
<b>BAB II TINJAUAN PUSTAKA</b>	

2.1 Data Mining dan Klasifikasi.....	8
2.2 Algoritma C 4.5 .....	10
2.3 Algoritma Naïve Bayes .....	11
2.4 Evaluasi Model Klasifikasi.....	12
2.5 Penelitian Terdahulu.....	13
2.6 Perbedaan Penelitian yang Dilakukan dengan Penelitian Terdahulu .	17

### **BAB III METODE PENELITIAN**

3.1 Objek Penelitian .....	18
3.2 Jenis Data .....	18
3.3 Metode Pengumpulan Data .....	18
3.3.1 Wawancara .....	18
3.3.2 Studi Pustaka .....	18
3.4 Metode Analisis Data .....	19
3.4.1 Data Selection.....	19
3.4.2 Pembersihan Data.....	20
3.4.3 Transformasi Data .....	21
3.4.4 Data Mining.....	24
3.4.5 Interpretasi dan Evaluasi .....	30
3.5 Implementasi Algoritma Data Mining.....	31
3.5.1 Algoritma C 4.5 .....	31
3.5.2 Algoritma Naïve Bayes .....	36

### **BAB IV HASIL PENELITIAN DAN PEMBAHASAN**

4.1 Pengujian Dataset dengan Rstudio .....	40
--	----

4.1.1 Instalasi Package Rstudio .....	40
4.1.2 Proses Import Dataset ke Rstudio.....	42
4.1.3 Pembagian Data Training dan Data Testing.....	43
4.2 Implementasi Dengan Algoritma C.45 .....	43
4.2.1 Percobaan ke-1 .....	43
4.2.2 Percobaan ke-2 .....	49
4.2.3 Percobaan ke-3 .....	55
4.2 Implementasi Dengan Algoritma Naïve Bayes .....	61
4.3.1 Percobaan ke-1 .....	61
4.3.2 Percobaan ke-2 .....	67
4.3.3 Percobaan ke-3 .....	73
4.4 Perbandingan Kinerja Algoritma C.45 dan Naïve Bayes .....	80

## **BAB V PENUTUP**

5.1 Kesimpulan.....	82
5.2 Saran .....	83

DAFTAR PUSTAKA

LAMPIRAN

## DAFTAR TABEL

Tabel	Halaman
2.1 Penjelasan tahapan dalam proses KDD.....	8
2.2 Ukuran evaluasi model klasifikasi .....	12
2.3 Perbandingan Penelitian Sebelumnya .....	16
3.1 Data penempatan PMI skema P2P pada BP3TKI Semarang .....	19
3.2 Penjelasan atribut dataset penempatan PMI.....	20
3.3 Ketentuan pengisian Atribut Status PMI.....	22
3.4 Atribut Dataset .....	23
3.5 Sampel Hasil Transformasi Data .....	24
3.6 Penjelasan Tahapan alur Data Mining .....	26
3.7 Penjelasan Tahapan alur Algoritma C.45.....	28
3.8 Penjelasan Tahapan alur Algoritma Naïve Bayes .....	30
3.9 Tabel perhitungan root node .....	33
3.10 Rule Penempatan PMI Berdasarkan Decision Tree .....	34
3.11 Contoh pengujian data testing pada rule Decision Tree.....	35
3.12 hasil perhitungan probabilitas class .....	37
3.13 Sample data testing untuk algoritma Naïve Bayes.....	39
3.14 Pengujian probabilitas algoritma Naïve Bayes .....	39
4.1 Tabel pengujian dataset dengan RStudio .....	40
5.2 Hasil Perbandingan Tingkat Akurasi dan Tingkat Error.....	80

## DAFTAR GAMBAR

Gambar	Halaman
2.1 Tahapan dalam proses KDD .....	8
3.1 Penjelasan penomoran NIK.....	21
3.2 Flowchart Data Mining Metode KDD .....	25
3.3 Flowchart Algoritma C.45 .....	27
3.4 Tahapan algoritma Naïve Bayes .....	29
3.5 Hasil Decision tree pada data penempatan PMI .....	34
4.1 Tampilan Dataset pada Rstudio .....	42
4.2 Tampilan Struktur Data dari Dataset.....	43
4.3 Hasil Pembagian Dataset.....	44
4.4 Decision Tree Algoritma C 4.5 Percobaan 1 .....	45
4.5 Hasil Rule Percobaan 1 .....	46
4.6 Hasil Prediksi Algoritma C 45 Percobaan 1 .....	47
4.7 Hasil Accuracy dan Error Rate Percobaan 1 .....	48
4.8 Hasil Pembagian Dataset Percobaan II .....	49
4.9 Decision Tree Algoritma C 4.5 Percobaan II.....	50
4.10 Hasil Rule Percobaan II .....	52
4.11 Hasil Prediksi Algoritma C 45 Percobaan II.....	53
4.12 Hasil Accurasi dan Error Rate Percobaan II .....	54
4.13 Hasil Pembagian Dataset Percobaan III.....	56
4.14 Decision Tree Algoritma C 4.5 Percobaan III .....	56

4.15 Hasil Rule Percobaan III .....	58
4.16 Hasil Prediksi Algoritma C 45 Percobaan III .....	59
4.17 Hasil Accuracy dan Error Rate Percobaan III.....	60
4.18 Hasil Probabilitas Naive Bayes Percobaan I.....	62
4.19 Hasil Prediksi Algoritma Naive Bayes Percobaan 1 .....	65
4.20 Hasil Akurasi Algoritma Naive Bayes Percobaan I.....	66
4.21 Hasil Probabilitas Naive Bayes Percobaan II.....	68
4.22 Hasil Prediksi Algoritma Naive Bayes Percobaan II .....	71
4.23 Hasil Akurasi Algoritma Naive Bayes Percobaan II.....	72
4.24 Hasil Probabilitas Naive Bayes Percobaan III .....	74
4.25 Hasil Prediksi Algoritma Naive Bayes Percobaan III.....	77
4.26 Hasil Akurasi Algoritma Naive Bayes Percobaan III .....	78
4.27 Chart Perbandingan Hasil Pengujian .....	79



## DAFTAR RUMUS

Rumus	Halaman
2.1 Formula Entropy .....	10
2.2 Information Gain .....	10
2.3 Split Info.....	11
2.4 Gain Ratio .....	11
2.5 Probabilitas bersyarat dalam teorema Bayes.....	11
2.6 Nilai maximum posteriori hypothesis. ....	11