

**IN
REVIEW**



**TELEGRAM BOT UNTUK PREDIKSI KEMAMPUAN BAYAR PESERTA
JKN MENGGUNAKAN ALGORITMA C4.5**

TUGAS AKHIR

Hersa Arnerangi Sandi
41517120112

UNIVERSITAS
MERCU BUANA
**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2020**

IN
REVIEW



**TELEGRAM BOT UNTUK PREDIKSI KEMAMPUAN BAYAR PESERTA
JKN MENGGUNAKAN ALGORITMA C4.5**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Hersa Arnerangi Sandi
41517120112

UNIVERSITAS
MERCU BUANA

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2020

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41517120112

Nama : Hersa Arnerangi Sandi

Judul Tugas Akhir : Telegram Bot Untuk Prediksi Kemampuan Bayar Peserta JKN Menggunakan Algoritma C4.5

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.



UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Hersa Amerangi Sandi
NIM : 41517120112
Judul Tugas Akhir : Telegram Bot Untuk Prediksi Kemampuan Bayar Peserta JKN Menggunakan Algoritma C4.5

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

UNIVERSITA
MERCU BUANA

Jakarta, 22 Februari 2020
METERAI TEMPEL
7095115725425414
6000
R. LAAGU BUDHAI
Hersa Amerangi Sandi

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Hersa Amerangi Sandi
 NIM : 41517120112
 Judul Tugas Akhir : Telegram Bot Untuk Prediksi Kemampuan Bayar Peserta JKN Menggunakan Algoritma C4.5

Menyatakan bahwa Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan v
		Jurnal Nasional Terakreditasi v	
		Jurnal International Tidak Bereputasi	Diterima
		Jurnal International Bereputasi	
Disubmit/dipublikasikan di :	Nama Jurnal	: JUST IT : Jurnal Sistem Informasi, Teknologi Informasi dan Komputer	
	ISSN	: 2089 - 0256	
2	Kertas Kerja, Merupakan material hasil penelitian sebagai kelengkapan Artikel Jurnal. Terdiri dari (minimal 4)	Literatur Review	[v]
		Hasil analisa & perancangan aplikasi	[v]
		Source code	[v]
		Data set	[v]
		Tahapan eksperimen	[v]
		Hasil eksperimen seluruhnya	[v]
3	HAKI Disubmit / Terdaftar	HKI	Diajukan
		Paten	Tercatat
		No & Tanggal Permohonan	:
		No & Tanggal Pencatatan	:

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 22 Februari 2020

 Hersa Amerangi Sandi

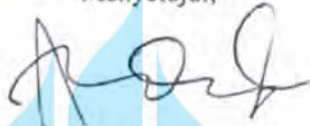
LEMBAR PERSETUJUAN

Nama Mahasiswa : Hersa Amerangi Sandi
NIM : 417120112
Judul Tugas Akhir : Telegram Bot Untuk Prediksi Kemampuan Bayar
Peserta JKN Menggunakan Algoritma C4.5

Tugas Akhir ini telah diperiksa dan disetujui

Jakarta, 22 Februari 2020

Menyetujui,



(Anis Cherid, SE, MTI)
Dosen Pembimbing



UNIVERSITAS
MERCU BUANA


LEMBAR PERSETUJUAN PENGUJI DAN PENGESAHAN

NIM : 41517120112
Nama : Hersa Arnerangi Sandi
Judul Tugas Akhir : Telegram Bot Untuk Prediksi Kemampuan Bayar Peserta JKN Menggunakan Algoritma C4.5

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 22 Februari 2020

Dewan Penguji:



(Drs. Achmad Kodar, MT)
Ketua Penguji



(Harni Kusniyati, M.Kom)
Anggota Penguji 1



(Umniy Salamah, S.Kom, MMSI)
Anggota Penguji 2

UNIVERSITAS
MERCU BUANA

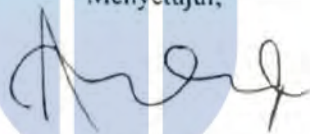
LEMBAR PENGESAHAN

NIM : 41517120112
Nama : Hersa Arnerangi Sandi
Judul Tugas Akhir : Telegram Bot Untuk Prediksi Kemampuan Bayar Peserta JKN Menggunakan Algoritma C4.5

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 22 Februari 2020

Menyetujui,



(Anis Cherid, SE, MTI)
Dosen Pembimbing

UNIVERSITAS

MENGETAHUI,
MERCU BUANA



(Diky Firdaus, S.Kom, MM)
Koord. Tugas Akhir Teknik Informatika



(Desi Ramavanti, S.Kom, MT)
Ka. Prodi Teknik Informatika

ABSTRACT

Name : Hersa Arnerangi Sandi
Student Number : 41517120112
Counsellor : Anis Cherid, SE, MTI
Title : Telegram Bot Untuk Prediksi Kemampuan Bayar Peserta JKN Menggunakan Algoritma C4.5

Main problem in Jaminan Kesehatan Nasional (JKN) program is the expenditure of health financing that is greater than the revenue collected, one of the causes is participants who are in arrears. The purpose of this study is to predict the ability to pay for participants who are expected to be one of the references in the process of registration or participation of class participants. To predict the ability to pay the authors use one prediction algorithm, the C4.5 algorithm. The data used in this study were collected from population census results stack up with payment data of 1324 rows. The prediction results are validated using a confusion matrix before being implemented into the telegram bot as an user interface. The accuracy generated by this algorithm for data models made more than 70% of this application will provide a pretty good prediction.

Keywords: *Algoritma C4.5, Prediction, Telegram Bot, Data Mining*



UNIVERSITAS
MERCU BUANA

KATA PENGANTAR

Puji syukur kita panjatkan kepada Allah SWT dan Nabi Muhammad SAW
Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak Oleh
karena itu, penulis mengucapkan terima kasih kepada:

1. Allah SWT
2. Orang Tua, Istri dan Keluarga Besar
3. Dosen beserta Jajaran Universitas Mercu Buana Jakarta
4. Pimpinan dan Rekan kerja di BPJS Kesehatan Cabang Bogor
5. Rekan-rekan Universitas Mercubuana
6. Seluruh pihak yang membantu penulis

Akhir kata, penulis berharap agar karya ini dapat bermanfaat di kemudian hari.

Jakarta, 22 Februari 2020
Penulis



UNIVERSITAS
MERCU BUANA

DAFTAR ISI

HALAMAN SAMPUL.....	0
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... ..	iii
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN	v
LEMBAR PERSETUJUAN PENGUJI	vi
LEMBAR PENGESAHAN	vii
ABSTRAK	viii
ABSTRACT	ix
KATA PENGANTAR.....	x
DAFTAR ISI.....	xi
NASKAH JURNAL	1
KERTAS KERJA.....	A
BAGIAN 1. LITERATUR REVIEW	B
BAGIAN 2 ANALISIS DAN PERANCANGAN.....	E
BAGIAN 3 SOURCE CODE	H
BAGIAN 4 DATASET	M
BAGIAN 5 TAHAPAN EKSPERIMEN	O
BAGIAN 6 HASIL SEMUA EKSPERIMEN.....	U
DAFTAR PUSTAKA	Y
LAMPIRAN.....	Z

TELEGRAM BOT UNTUK PREDIKSI KEMAMPUAN BAYAR PESERTA JKN MENGGUNAKAN ALGORITMA C4.5

Hersa Arnerangi Sandi¹, Anis Cherid²

Fakultas Ilmu Komputer Universitas Mecu Buana
hersa87@ymail.com¹, anis.cherid@mercubuana.ac.id²

Abstrak

Permasalahan dalam program Jaminan Kesehatan Nasional (JKN) adalah pengeluaran pembiayaan pelayanan kesehatan yang lebih besar daripada iuran yang dikumpulkan, salah satu penyebabnya adalah peserta yang menunggak iuran. Tujuan dari penelitian ini adalah untuk memprediksi kemampuan bayar peserta sehingga diharapkan dapat menjadi salah satu acuan dalam proses pendaftaran atau perpindahan kelas peserta. Untuk memprediksi kemampuan bayar tersebut penulis menggunakan salah satu algoritma prediksi yaitu algoritma C4.5. Dataset yang digunakan pada penelitian ini dikumpulkan dari hasil sensus penduduk dibandingkan dengan data pembayaran sebanyak 1324 baris. Hasil prediksi divalidasi menggunakan confusion matrix sebelum diimplementasikan ke dalam telegram bot sebagai antarmuka. Akurasi yang dihasilkan algoritma ini terhadap model data yang dibuat sebesar 77 % sehingga aplikasi ini akan memberikan prediksi yang cukup baik.

Kata Kunci: Algoritma C4.5, Prediksi, Telegram Bot, Data Mining

Abstract

Main problem in Jaminan Kesehatan Nasional (JKN) program is the expenditure of health financing that is greater than the revenue collected, one of the causes is participants who are in arrears. The purpose of this study is to predict the ability to pay for participants who are expected to be one of the references in the process of registration or participation of class participants. To predict the ability to pay the authors use one prediction algorithm, the C4.5 algorithm. The data used in this study were collected from population census results stack up with payment data of 1324 rows. The prediction results are validated using a confusion matrix before being implemented into the telegram bot as an user interface. The accuracy generated by this algorithm for data models is 77% so this application will provide a pretty good prediction.

Keywords: Alogritma C4.5, Prediction, Telegram Bot, Data Mining

1. Pendahuluan

Jaminan Kesehatan Nasional (JKN) merupakan bagian dari Sistem Jaminan Sosial Nasional (SJSN) yang diselenggarakan dengan menggunakan mekanisme asuransi kesehatan sosial yang bersifat wajib (*mandatory*) berdasarkan Undang-Undang Nomor 40 Tahun 2004 tentang SJSN dengan tujuan untuk memenuhi kebutuhan dasar kesehatan masyarakat yang layak yang

diberikan kepada setiap orang yang telah membayar iuran atau iurannya dibayar oleh Pemerintah. Peserta JKN adalah penduduk Indonesia baik yang bekerja sebagai karyawan, wiraswasta, dan pekerjaan lainnya yang tidak dapat dipastikan gajinya. Saat ini pemilihan kelas untuk Peserta Bukan Penerima Upah / Mandiri dibebaskan, tidak seperti pegawai swasta atau PNS yang dapat diukur kelas maksimal sesuai gaji atau golongannya. Sehingga banyak peserta yang

menunggak iuran dikarenakan merasa berat akibat salah memilih kelas yang tidak sesuai kemampuan ekonominya, atau sengaja memilih kelas paling tinggi kemudian hanya membayar pada saat sakit saja.

Berdasarkan uraian diatas, perlu dilakukan analisa untuk memberikan solusi terhadap permasalahan dalam menganalisa dengan menentukan prediksi kelancaran pembayaran iuran sesuai kelas peserta menggunakan pendekatan *data driven method* yaitu suatu pengambilan keputusan untuk pengalisan lebih lanjut didasari kepada data peserta yang telah dibuat datasetnya. Algoritma yang digunakan untuk pengklasifikasian dan prediksi data adalah Algoritma C4.5

Algoritma C4.5 banyak digunakan dan diimplementasikan dalam suatu pengambilan keputusan karena kelebihan dari Algoritma C4.5 dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang kosong, membuat aturan yang mudah untuk diimplentasikan dan performanya merupakan salah satu yang tercepat dibandingkan dengan algoritma lain dan dapat memprediksi kelas objek yang tidak diketahui. Kemudian penulis membuat Telegram Bot dengan Python sebagai pemrosesnya untuk mempermudah penggunaan hasil penelitian ini.

Tujuan dari penelitian ini untuk memprediksi kemampuan bayar peserta sehingga diharapkan dapat menjadi salah satu acuan dalam proses pendaftaran atau perpindahan kelas peserta. Kemudian manfaat dari penelitan ini bagi organisasi adalah untuk melakukan tindakan pencegahan menunggaknya peserta mandiri, dan bagi peserta adalah peserta dapat membayar iuran sesuai dengan kemampuannya sehingga tidak terbebani.

Penulis menggunakan beberapa jurnal penelitian terdahulu yang hamper sama sebagai salah satu acuan penulis dalam melakukan penelitian, sehingga dapat mempermudah dalam mengkaji penelitian. Berikut merupakan jurnal penelitian terdahulu yang terkait dengan penelitian ini.

TABEL 1
PENELITIAN TERDAHULU

Nama Peneliti	Judul Penelitian	Intisari
---------------	------------------	----------

TB Santoso	Analisa Dan Penerapan Metode C4.5 Untuk Prediksi Loyalitas Pelanggan	Berdasarkan analisa penggunaan data mining dengan algoritma C4.5 dapat digunakan pada data set pelanggan kedalam kegiatan manajemen strategi sehingga dapat menahan selama mungkin pelanggannya dengan baik (Santoso, 2014).
DH Kamagi, S Hansun	Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa	Data mining dengan algoritma C4.5 dapat diimplementasikan untuk memprediksi tingkat kelulusan mahasiswa dengan empat kategori yaitu lulus cepat, lulus tepat, lulus terlambat dan <i>drop out</i> (Kamagi & Hansun, 2014).
Mohamad Efendi Lasulika	Komparasi Naïve Bayes, Support Vector Machine Dan Knearest Neighbor Untuk Mengetahu	Berdasarkan hasil komparasi antara algoritma klasifikasi data mining, maka didapatkan hasil analisa bahwa metode Naïve Bayes merupakan algoritma yang

	i Akurasi Tertinggi Pada Prediksi Kelancaran Pembayaran n TV Kabel	akurat menghasilkan accuracy yaitu 96 %, dan metode ini juga sangat dominan terhadap metode lain (Lasulika, 2019).
	Implementasi Data Mining dengan Algoritma C4.5 pada Penjualan Obat	Algoritma C4.5 ini dapat diterapkan di objek-objek data set lainnya, baik dataset yang dikelola oleh pemerintah, BUMN ataupun Swasta agar masyarakat dapat mengetahui berita berdasarkan data yang ada bukan lagi dari prediksi orang perorang. (Avrizal & Dwiasnati, 2019)
Muhammad Fauzul Arifin, Devi Fitriannah	Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada	Penerapan algoritma klasifikasi C4.5 akan dapat diimplementasikan pada rekomendasi penerimaan mitra penjualan di PT. Atria Artha Persada, dilihat dari tingkat accuracy yang mencapai 96.26 % dan recall 71.43% (Arifin & Fitriannah,

		2018).
--	--	--------

2. Metode Penelitian

Metode penelitian yang digunakan dalam penerapan algoritma C4.5 untuk prediksi dan klasifikasi pembayaran peserta JKN secara garis besar menggunakan rancangan penelitian pada gambar 1.



Gambar 1. Flowchart Rancangan Proses Penelitian

Pengumpulan Data

Proses yang dilakukan untuk mengumpulkan data yang akan digunakan dalam proses pembuatan dataset.

Seleksi Data

Diperlukan seleksi data untuk memilih data yang digunakan untuk menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana prediksi akan dilakukan.

Transformasi Data

Transformasi data adalah suatu proses pengolahan data dan perubahan data kedalam format atau bentuk yang sesuai, sehingga dapat diproses dengan perhitungan algoritma C4.5. Transformasi data dapat dikatakan sebagai proses transformasi atau penggabungan data- kedalam bentuk yang diinginkan / sesuai untuk dilakukan proses *data mining* melalui operasi *summary* atau *aggregation*.

Perhitungan Entropy dan Gain Ratio

Perhitungan seluruh atribut entropy menggunakan rumus (1) dan *gain ratio* menggunakan rumus (2) untuk mencari dan mengetahui *gain ratio* tertinggi yang akan dijadikan simpul akar pada proses pembuatan

keputusan (Quinlan, 1993). Untuk menghitung entropi menggunakan rumus berikut :

$$\text{Entropi (S)} = \sum_{j=1}^k - p_j \log_2 p_j \quad (1)$$

Dimana:

S = himpunan (dataset) kasus

k = banyaknya partisi S

p_j = probabilitas yang di dapat dari Sum(Ya) dibagi Total Kasus.

Kemudian untuk Gain Ratio dapat dihitung dengan rumus :

$$\text{Gain Ratio(a)} = \frac{\text{gain(a)}}{\text{split(a)}} \quad (2)$$

Keterangan:

a = atribut.

gain(a) = information gain pada atribut a / rumus (3)

Split(a) = split information pada atribut a / rumus (4)

Split menyatakan entropy atau informasi potensial yang bisa dihitung menggunakan rumus :

$$\text{Split(S,A)} = \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (3)$$

Keterangan:

S = ruang (data) sampel training

A = atribut.

S_i = jumlah sample untuk atribut i

Untuk menghitung gain dapat menggunakan rumus :

$$\text{Gain(A)} = \text{Entropi(S)} - \sum_{i=1}^k \frac{|S_i|}{|S|} \times$$

Entropi(S_i) (4)

Keterangan:

S = ruang (data) sampel training

A = atribut.

$|S_i|$ = jumlah sample untuk nilai V.

$|S|$ = jumlah seluruh sample data.

Entropi(S_i) = entropy untuk sample yang memiliki nilai i

Pohon keputusan (*Decision Tree*)

Pohon keputusan adalah hasil dari proses perhitungan *entropy* dan *information gain*, setelah dilakukan perhitungan berulang-ulang sampai semua atribut mempunyai kelas dan tidak dapat dilakukan proses perhitungan.

Validasi dan Pengujian

Validasi dan pengujian yang dilakukan adalah untuk memastikan bahwa hasil prediksi sudah dilakukan dengan benar dan sesuai dengan pembahasan. Proses ini akan dilakukan dengan metode *confusion matrix*. *Confusion matrix* adalah model yang akan membentuk matriks yang terdiri dari *true positive* dan *true negative*.

TABEL 2
CONFUSION MATRIX

	Prediksi Lunas	Prediksi Menunggak
Aktual Lunas	TP	FN
Aktual Menunggak	FP	TN

Keterangan :

True Positive (TP) : seharusnya positif dan diprediksi positif.

False Negative (FN) : seharusnya positif dan diprediksi negatif.

True Negative (TN) : seharusnya negatif dan diprediksi negatif.

False Positive (FP) : seharusnya negatif dan diprediksi positif.

Setelah membuat tabel diatas maka dapat dihitung akurasi, presisi dan recall dengan rumus berikut

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (7)$$

Evaluasi dan Presentasi

Pada proses ini akan dilakukan evaluasi sesuai hasil dari validasi dan pengujian. Kemudian dilakukan pembuatan bot telegram dan python

sebagai pemroses datanya untuk prediksi agar hasil analisa yang dilakukan pada penelitian ini mudah diaplikasikan atau digunakan oleh pengguna.

3. Hasil dan Pembahasan

Pengumpulan Data

Data yang akan digunakan berjumlah 1324 baris yang dikumpulkan dari data pembangunan keluarga hasil dari sensus lapangan dan disandingkan dengan data peserta menunggak bulan april sampai september dengan rincian atribut seperti tabel 2 berikut.

TABEL 3
DATA AWAL

PK19	PK20	PK21	PK22	PK23	PK26	PK27	PK28	KELAS	Decision
2	2	3	9	4	1	30	6	3	Menunggak
1	2	9	9	4	1	160	6	3	Lancar
1	2	3	8	4	1	90	5	2	Menunggak
1	1	7	8	4	1	28	3	3	Menunggak
2	1	5	8	4	4	99	4	3	Menunggak
2	1	3	8	4	1	52	3	3	Lancar

Seleksi Data

Data awal kemudian diseleksi dan difokuskan pada 6 atribut dengan rincian seperti tabel 3 berikut.

TABEL 4
SELEKSI DATA

sumber_air	kepemilikan	luas_rumah	jumlah_huni	kelas	Decision
1	1	150	2	3	Menunggak
1	1	53	4	2	Menunggak
2	3	180	3	1	Menunggak
1	3	34	5	1	Menunggak
1	1	75	3	2	Menunggak
2	1	93	3	1	Menunggak

Transformasi Data

Hasil seleksi data ditransformasikan supaya dapat diproses menggunakan algoritma c4.5. Data yang ditransformasikan adalah sumber air, dan kepemilikan karena masih berupa id pada referensi datanya. Pada atribut luas rumah sesuai usulan dari organisasi masyarakat penilai Indonesia diklasifikasikan sebagai berikut :

- Rumah Sangat Sederhana (RSS) luas <36
- Rumah Sederhana (RS) 36-54m²
- Rumah Kelas Menengah luas 54-120m²
- Rumah Kelas Mewah luas > 120m²

Kemudian dibuat atribut baru yaitu kjm (kebutuhan jiwa permeter) berdasarkan perhitungan kebutuhan jiwa per meter diklasifikasikan sesuai Keputusan Menteri Permukiman dan Prasarana Wilayah nomor 403/KPTS/M/2002 yaitu kebutuhan luas ruang/jiwa minimum adalah 9 m²/jiwa maka dibuat kategori kurang ideal, ideal, dan sangat ideal. Hasil transformasi tabel 2 dapat dilihat pada tabel 4 berikut.

TABEL 5
TRANSFORMASI DATA

sumber_air	kepemilikan	luas_rumah	kjm	kelas	ket
Ledeng	Sendiri	Menengah	ideal	3	lancar
Ledeng	Sendiri	Mewah	Sangat Ideal	2	lancar
Ledeng	Sendiri	Menengah	Sangat Ideal	3	lancar
Ledeng	Sewa	Menengah	ideal	3	lancar
Ledeng	Sendiri	Mewah	Sangat Ideal	2	lancar
Ledeng	Menumpang	Menengah	ideal	3	lancar
Ledeng	Sendiri	RS	Sangat Ideal	2	lancar

Penghitungan Entropy dan Gain Ratio

Setelah transformasi data, perlu dilakukan penghitungan entropi dan gain ratio masing-masing atribut agar dapat dipilih atribut yang mempunyai gain ratio tertinggi sehingga dapat menghasilkan tree yang kecil atau dengan sedikit node. Untuk memudahkan perhitungan dibuat dalam rumus excel yang hasilnya dapat dilihat pada tabel 5 berikut.

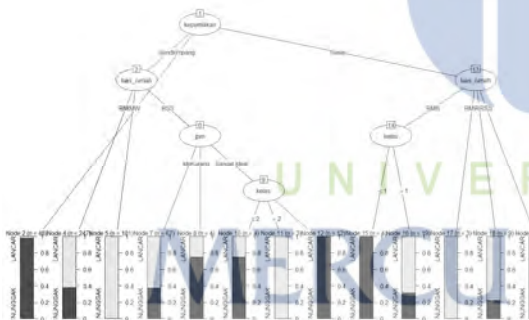
TABEL 6
PENGHITUNGAN GAIN RATIO

ATRIBUT	DATA	JML KASUS	Nunggak (S1)	Lancar (S2)
TOTAL		1324	509	551
SUMBER AIR	Ledeng	1208	448	760
	Sumur	116	61	55
KEPEMILIKAN	Kontrak	106	63	43
	Menumpang	108	101	2
	Milik Sendiri	1115	343	772
KELAS	Kelas 1	308	77	231
	Kelas 2	360	135	225
	Kelas 3	656	297	359
KEB JIWA PERMETER	ideal	845	375	470
	Kurang	79	67	12
	Sangat Ideal	400	67	333
LUAS RUMAH	Menengah	742	318	424
	Mewah	266	1	265
	RS	234	109	125
	RSS	82	82	0

ENTROPY	GAIN INFO	SPLIT INFO	GAIN RATIO
1,056562268			
	0,10113652	0,428452724	0,236050593
0,951330842			
0,998069255			
	0,21796717	0,786984736	0,276964928
0,974165552			
0,138157229			
0,890410912			
	0,11605119	1,502264836	0,07725082
0,811278124			
0,954434003			
0,993546896			
	0,190541112	1,177852423	0,161769936
0,990863141			
0,614571561			
0,65195376			
	0,321105504	1,623811224	0,19774805
0,985228136			
0,035696475			
0,996624861			
0			

Pohon Keputusan

Setelah menemukan gain tertinggi akan memulai membuat tree. Pada tabel 2 dapat dilihat bahwa paling tinggi gain rasionya adalah atribut kepemilikan. Maka node pertama / akar dari pohon akan dimulai dari kepemilikan. Kemudian setelah itu ulangi penghitungan entropy dan gain rasionya untuk atribut selanjutnya sehingga dapat dihasilkan pohon keputusannya.



Gambar 2. Pohon Keputusan

Validasi dan Pengujian

Perlu dilakukan pengujian menggunakan *confusion matrix* agar mengetahui tingkat ketepatan prediksinya. Data yang diuji sejumlah 103 baris kemudian hasilnya dimasukkan ke dalam table confusion matrix sehingga dapat dihitung akurasi, recall dan precisionnya menggunakan rumus (5), (6), dan (7)

TABEL 7
PENGHITUNGAN *CONFUSION MATRIX*

N = 103	Prediksi Lunas	Prediksi Menunggak
Aktual Lunas	34	13
Aktual Menunggak	18	38

$$Accuracy = \frac{34+3}{34+38+18+13} = 0,774$$

$$Recall = \frac{34}{34+13} = 0,723$$

$$Precision = \frac{34}{34+18} = 0,653$$

Selain menggunakan rumus diatas dataset yang digunakan juga telah diuji menggunakan algoritma c4.5 yang tersedia di RStudio untuk memastikan keaakuratannya, hasil pengujiannya dapat dilihat pada gambar 3 berikut.

```

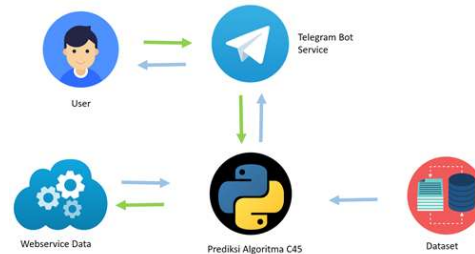
> c45_model <- J48(as.factor(Decision)~., dataTrain)
> dataTest.pred <- predict(c45_model, newdata = dataTest)
> cMatrix <- table(dataTest.pred, dataTest$Decision)
> summary(c45_model)

*** Summary ***
Correctly Classified Instances      400          75.6144 %
Incorrectly Classified Instances    129          24.3856 %
Kappa statistic                    0.4179
Mean absolute error                 0.3092
Root mean squared error             0.3932
Relative absolute error             65.3699 %
Root relative squared error        80.8603 %
Total number of instances          529
    
```

Gambar 3. Hasil pengujian RStudio

Evaluasi dan Presentasi

Akurasi sudah cukup baik yaitu diatas 70% sehingga dapat dilanjutkan ke tahap presentasi dengan membuat antarmuka aplikasi menggunakan Bot Telegram yang diberi nama Telegram bot. Bot Telegram adalah layanan chat yang bisa dapat di program agar dapat menjalankan perintah sesuai kebutuhan kita. Alur kerja bot telegram yang dibuat dapat dilihat pada gambar 4 berikut.



Gambar 4. Arsitektur Aplikasi

Untuk membuat telegram bot, pertama-tama harus mendaftar akun telegram kemudian mengirim *chat* ke @BotFather, setelah menentukan nama dan deskripsinya, bot tersebut akan mengirimkan token yang nantinya akan digunakan untuk mengambil data *chat* yang masuk ke bot yang telah dibuat menggunakan metode *getUpdates* yang telah disediakan oleh Telegram API. Salah satu kelebihan python adalah memiliki koleksi kepustakaan (*library*) yang banyak, yaitu telah tersedia modul-modul siap pakai untuk berbagai keperluan (Enterprise, 2018). Untuk pengambilan data *chat* di telegram penulis menggunakan modul python-telegram-bot, cara menginstal modul tersebut dengan menjalankan perintah berikut di *command line*:

```
python -m pip install python-telegram-bot
```

Setelah terinstal cukup mengimport modul tersebut dan memasukkan token yang telah didapat sebelumnya dengan contoh sebagai berikut :

```
from telegram.ext import Updater,
CommandHandler

def hello(update, context):
    update.message.reply_text(
        'Hello
    {}'.format(update.message.from_user.fi
rst_name))

updater = Updater('DIISI TOKEN
TELEGRAM', use_context=True)
updater.dispatcher.add_handler(Command
Handler('hello', hello))
updater.start_polling()
updater.idle()
```

Setelah bisa mendapatkan dan mengirim chat melalui bot, selanjutnya adalah membuat proses prediksinya menggunakan algoritma C4.5. untuk proses ini digunakan framework Chefboost dan Pandas (*Python Data Analysis Library*). Cara menginstal framework tersebut pada python menggunakan *command line* berikut :

```
python -m pip install pandas
```

Framework tersebut dapat digunakan dengan mengimport ke script telegram yang dibuat sebelumnya. Untuk membuat model dari dataset yang telah dibuat untuk melakukan prediksi menggunakan kode sebagai berikut :

```
import Chefboost as chef
import pandas as pd

df =
pd.read_csv("dataset/data_jurnal.csv")

config = {'algorithm': 'C4.5'}
model = chef.fit(df.copy(), config)
chef.save_model(model,
"prediksi_lancar.pkl")
```

Model tersebut akan tersimpan di direktori dimana kita menyimpan script code python dan dapat digunakan untuk memprediksi dengan contoh sebagai berikut :

```
datanya = ['Ledeng', 'Sendiri', 'RSS',
'Ideal', '2']
model =
chef.load_model("prediksi_lancar.pkl")
prediksi = chef.predict(model, datanya)
```

Aplikasi sudah dapat melakukan prediksi terhadap *chat* yang dikirim oleh pengguna. Selanjutnya dilakukan proses pengujian langsung ke telegram bot dengan menggunakan salah satu data latih.

Proses pengujiannya pertama user akan mengirimkan pesan melalui telegram ke telegram bot. Kemudian bot akan memberikan response sesuai inputan dari user yang diproses melalui script python yang telah terintegrasi dengan webservice data dan dataset yang telah dibuat sebelumnya. Tampilannya awal Telegram bot akan menampilkan pertanyaan untuk memilih layanan yang akan digunakan, yaitu prediksi kelancaran bayar peserta, untuk lebih jelasnya dapat dilihat pada gambar berikut.



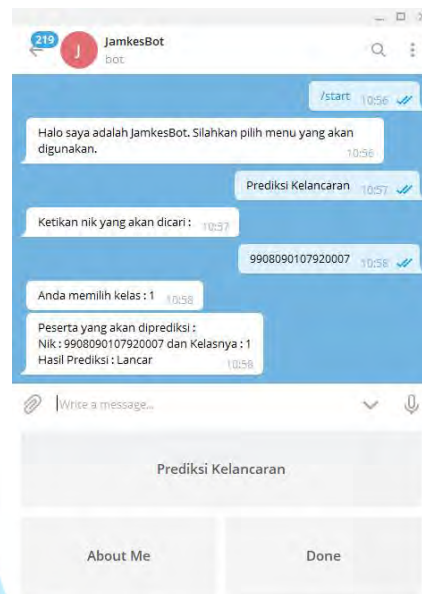
Gambar 5. Tampilan awal Aplikasi Bot Telegram

Setelah memilih menu Prediksi Kelancaran, bot akan meminta pengguna memasukan NIK orang yang akan dianalisa datanya seperti gambar berikut



Gambar 6. Input NIK Aplikasi Bot Telegram

Kemudian bot akan menanyakan kelas asuransi yang akan diambil, pengguna dapat memilih salah satu diantara 3 kelas yang tersedia. Setelah memilih kelasnya bot akan mengambil data ke webservice berdasarkan NIK yang telah diinputkan dan meprediksi data tersebut menggunakan dataset yang telah dibuat sehingga dapat memprediksi kelancaran bayarnya sesuai gambar dibawah



Gambar 7. Input Kelas dan Prediksi Aplikasi Bot Telegram

Hasil prediksinya membutuhkan waktu 3-8 detik saat pengujian dan ditampilkan dalam teks keterangan lancar atau menunggak.

4. Simpulan dan Saran

Dari uraian pada bab-bab yang sudah dibahas sebelumnya dapat ditarik simpulan sebagai berikut:

a. Untuk memprediksi kelancaran bayar peserta diperlukan atribut berikut ini :

- 1) Kepemilikan rumah
- 2) Jenis air yang digunakan
- 3) Luas rumah
- 4) Jumlah penghuni permeter
- 5) Kelas rawat

b. Algoritma C4.5 dalam memprediksi memprediksi kelancaran bayar peserta cukup baik performanya karena tergolong cepat saat diterapkan pada aplikasi Telegram bot dan akurasiya cukup baik yaitu 77%.

c. Dataset yang digunakan masih belum optimal sehingga kemungkinan besar dapat diganti dengan dataset lain agar mendapatkan akurasi prediksi yang lebih baik.

Daftar Pustaka

- Arifin, M. F., & Fitriannah, D. (2018). Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada. *IncomTech*, 87-102.
- Avrizal, R., & Dwiasnati, S. (2019). Implementasi Data Mining dengan Algoritma C4.5 pada Penjualan Obat. *FORMAT*, 77-84.
- Enterprise, J. (2018). *Otodidak Pemrograman Python*. Jakarta: PT Elex Media Komputindo.
- Kamagi, D. H., & Hansun, S. (2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *ULTIMATICS*, 15-20.
- Kusrini, & luthfi, E. t. (2009). *Algoritma Data Mining*. Yogyakarta: Penerbit Andi.
- Lasulika, M. E. (2019). Komparasi Naïve Bayes, Support Vector Machine Dan Knearest Neighbor Untuk Mengetahui Akurasi Tertinggi Pada Prediksi Kelancaran Pembayaran TV Kabel. *ILKOM Jurnal Ilmiah*, 11-16.
- Lusinia, S. A. (2014). Algoritma C4.5 Dalam Menganalisa Kelayakan Kredit(Studi Kasus Di Koperasi Pegawai Republik Indonesia (KP-RI) Lengayang Pesisir Selatan. *Jurnal KomTekInfo Fakultas Ilmu Komputer*, 6-10.
- Modrzyk, N. (2014). *Building Telegram Bots*. California: Apress Media.
- P.Premchand, M. M. (2014). Performance Analysis of Decision Tree Classifier. *International Journal of Computer Science Trends and Technology*, 62-70.
- Quinlan, J. R. (1993). *Programs for Machine Learning*. San Mateo: Morgan Kauffman.
- Rani, L. N. (2016). Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit. *JURNAL INOVTEK POLBENG*, 126-132.
- Santoso, T. B. (2014). Analisa Dan Penerapan Metode C4.5 Untuk Prediksi Loyalitas Pelanggan. *Jurnal Ilmiah LIMIT'S*, 33-36.
- Widayu, D. H., Nasution, S. D., Silalahi, N., & Mesran. (2017). Data Mining untuk memprediksi jenis transaksi nasabah pada koperasi simpan pinjam. *MEDIA INFORMATIKA BUDIDARMA*, 32-37.