

# Penerapan Metode Klasifikasi C4.5 dalam Pembuatan Website Identifikasi untuk Prediksi Kredibilitas Akun pada Media Sosial Instagram

Yonas Christianto, Rolly Intan, Rudy Adipranata

Program Studi Informatika, Fakultas Teknologi Industri, Universitas Kristen Petra

Jln. Siwalankerto 121 – 131 Surabaya 60236

Telp. (031)-2983455, Fax. (031)-8417658

yonaschristianto@yahoo.co.id, rintan@petra.ac.id, rudya@petra.ac.id

## ABSTRAK

Instagram adalah salah satu platform sosial media yang paling besar saat ini. Sebagai platform yang besar, semakin banyak juga pihak-pihak yang memanfaatkan platform ini untuk hal-hal yang tidak bertanggung jawab seperti membuat akun palsu. Permasalahan yang ingin diselesaikan oleh penulis adalah dengan memanfaatkan sebuah website yang berfungsi untuk mencari informasi tentang kredibilitas sebuah akun Instagram, agar pengguna dapat berinteraksi dengan aman dan nyaman di media sosial Instagram. Metode yang digunakan untuk melakukan prediksi adalah klasifikasi C4.5 yang menghasilkan decision rules. Decision rules inilah yang akan digunakan oleh sistem untuk memprediksi kredibilitas akun Instagram. Berdasarkan pengujian yang telah dilakukan, web yang dibuat telah dapat digunakan untuk menentukan kredibilitas akun Instagram seseorang dan hasil dari uji metode klasifikasi mampu mencapai nilai akurasi 97,07%.

**Kata Kunci:** klasifikasi C4.5, akun Instagram, *Confusion Matrix*

## ABSTRACT

*Instagram is one of the biggest social media platforms nowadays. As one of the biggest platforms, there's also a lot of people who makes this platform become unhealthy social media environment, by making fake or spam accounts. This is the problem that the author is trying to solve. By developing a website that people can use to gain information about the credibility of an Instagram account, so users can interact with the target accounts safely and comfortably. The method used to predict the credibility is C4.5 classification which produce decision rules. This decision rules will be used to predict the credibility of an Instagram account. Based on the test that have been carried out, the website can be used to determine the credibility of an Instagram account and the result of the classification method reached to 97,07%.*

**Keywords:** C4.5 classification, Instagram account, *Confusion Matrix*

## 1. PENDAHULUAN

Instagram merupakan media sosial dimana para penggunanya dapat dengan mudah berpartisipasi dalam berbagi gambar, menulis, dan mencipta sesuatu, seperti blogging, berkomunikasi secara virtual, dan memamerkan karya. Seiring dengan berjalannya waktu, peranan *Instagram* semakin berkembang. Tidak hanya sebagai sarana menjalin hubungan komunikasi di dunia online, tetapi Instagram dapat digunakan juga sebagai sarana untuk berbisnis,

berpolitik, dan bidang pendidikan. Salah satu sarana pemakaian media sosial Instagram saat ini adalah untuk berbisnis, yaitu pemasaran screen to face atau yang dikenal sebagai internet marketing [10]. Dalam dunia bisnis, persaingan di media sosial terkadang dapat mendesak pelaku bisnis untuk menghalalkan berbagai cara dalam mendapatkan keuntungan, termasuk melakukan penipuan di antaranya pemakaian akun palsu.

Berdasarkan permasalahan diatas, usaha untuk memerangi eksploitasi penggunaan akun pada Instagram dinilai perlu. Adanya sebuah web identifikasi yang dapat memberi keterangan mengenai kredibilitas akan suatu profil atau akun di Instagram, akan memberikan kenyamanan dan meningkatkan keamanan saat berjejaring sosial. Kriteria yang menjadi penentu kredibilitas sebuah akun memiliki banyak variabel. Oleh karena itu, sebuah web identifikasi akan mempermudah pengguna untuk mendapatkan keterangan mengenai kredibilitas akun yang pengguna temui, maupun akun mereka sendiri.

Web identifikasi ini bekerja dengan cara melakukan prediksi akun media sosial Instagram yang telah user inputkan dengan menggunakan decision rules yang sudah didapatkan dari hasil klasifikasi pada dataset "Instagram Fake Spammer Genuine Accounts". Metode yang digunakan untuk proses klasifikasi dan menentukan decision rules adalah algoritma C4.5. Algoritma C4.5 ini akan diterapkan ke dataset untuk membangun decision rules-nya. Input dari user yang berupa *username* akun tertentu akan diekstraksi atribut-atributnya menggunakan Instagram Scraper. Atribut-atribut hasil input akan diprediksi menggunakan *decision rules* yang telah didapat sehingga dapat diidentifikasi apakah akun tersebut adalah palsu atau akun asli.

Aplikasi identifikasi semacam ini telah terdapat di aplikasi mobile salah satunya adalah Truecaller. Truecaller adalah sebuah aplikasi mobile dengan lebih dari 200 juta pengguna di dunia [1]. Fitur-fitur yang tersedia di Truecaller meliputi identifikasi mengenai siapa yang menghubungi kita, blokir telepon, merekam percakapan telepon, dan sebagainya. Aplikasi identifikasi yang digunakan hanya berlaku pada telepon seluler saja. Cara kerjanya dengan menyimpan laporan dari user berdasarkan nomor telepon yang pernah menghubunginya. Apabila nomor yang banyak mendapat laporan penipuan menghubungi user, maka user tersebut akan diberi peringatan bahwa itu adalah nomor penipuan.

Berbeda dengan Truecaller, web identifikasi yang akan digunakan dalam penelitian ini terfokus pada atribut yang dimiliki oleh sebuah akun dan dibandingkan dengan kumpulan rules yang telah dihasilkan oleh proses klasifikasi. Tidak berdasarkan pada pengalaman dan laporan sesama pengguna. Metode pengujian yang digunakan juga berbeda, skenario yang akan digunakan untuk

pengujian nantinya berupa pengujian kepada beberapa kelompok data dengan jumlah perbandingan data latih dan data uji yang berbeda. Setiap pengujian tersebut akan menghasilkan nilai akurasi dan nilai akurasi terbaik yang akan digunakan untuk sistem website.

## 2. LANDASAN TEORI

### 2.1 Algoritma C4.5

Algoritma C4.5 adalah sebuah algoritma induksi decision tree yaitu ID3 dalam data mining, yang dikembangkan oleh J. Ross Quinlan. Pada dasarnya, konsep dari algoritma ini adalah membentuk suatu decision tree, berdasarkan atribut yang memiliki nilai prioritas tertinggi atau memiliki nilai gain tertinggi dari entropy atribut-atributnya. Nilai Gain yang tertinggi yang akan dijadikan poros atribut klasifikasi. Dari poros tersebut, akan diperluas secara rekursif cabang-cabang decision tree-nya [9]. Untuk membuat *decision tree* terdapat empat langkah yang dilakukan, diantaranya memilih atribut sebagai akar, membuat cabang untuk tiap nilai, membagi kasus ke dalam cabang, dan mengulangi proses yang sama dalam tiap cabang hingga tiap cabang memiliki kelas yang sama.

Setelah langkah-langkah diatas, lakukan pencarian nilai entropy dan gain.

$$Entropy(S) = \sum_{i=1}^n -P_i * Log_2 P_i \quad (1)$$

Perhitungan nilai *entropy* berdasarkan persamaan 1 digunakan untuk menentukan *heterogenity* dari kumpulan data.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Nilai *gain* pada persamaan 2 digunakan untuk menentukan *root* dari *decision tree*.

Alasan mengapa algoritma C4.5 banyak digunakan karena memiliki kelebihan mampu menghasilkan *decision tree* yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut yang bertipe diskret, dan dapat menangani atribut bertipe diskret dan numerik [8]. Algoritma C4.5 juga dinilai sangat kuat apabila digunakan untuk mengklasifikasi data yang *missing values*, data yang bersifat berulang atau *continous*, dan *pruning* [7].

### 2.2 Pandas Library pada Python

*Pandas* adalah singkatan dari “*Python Data Analysis Library*” yang merupakan sebuah *library* pada bahasa pemrograman *Python* untuk melakukan analisis data. *Pandas* dinilai sangat membantu untuk kegiatan analisis data, dan menjadi salah satu *library* pada *python* yang paling banyak digunakan. Penggunaan *Pandas Library* jauh lebih mudah jika dibandingkan dengan menggunakan *list*, *dictionaries through loops*, atau *list comprehension* pada *Python*. Cara kerja *library* ini adalah membawa data dalam format CSV, TSV, ataupun SQL Database dan menjadikannya sebuah obyek *Python* dengan baris dan kolom bernama *dataframe*. Obyek ini bentuk dan kegunaanya mirip dengan tabel-tabel dalam *statistical software* seperti *Microsoft Excel* atau *SPSS* [5].

### 2.3 Chefboost Library pada Python

Chefboost adalah sebuah *library* pada *python* yang menyediakan layanan *framework* ringan dan mudah untuk mencari *decision tree*. Metode pencarian *decision tree* yang disediakan oleh *chefboost* beraneka ragam mulai dari *gradient boosting*, *random forest*, *adaboost*, *ID3*, *C4.5*, *CART*, *CHAID*, dan *regression tree*. Pada

penelitian ini akan digunakan algoritma C4.5 dari *Chefboost* untuk melakukan penelitian terhadap dataset [12].

### 2.4 Web Scraping

Web Scraping adalah sebuah alat untuk mengambil data dari sebuah web, misalnya mengambil data kumpulan suatu kontak pada website. Keunggulan dari web scraping di antaranya adalah dapat mengumpulkan data dengan lebih cepat, automasi juga dapat diterapkan menggunakan web scraping, sehingga dari efisiensi tersebut, proses analisa data juga dapat dilakukan lebih komprehensif. Hasil ekstrasi dari web scraping dapat disimpan dalam database, maupun file. Web scraping dapat dilakukan dengan menggunakan web scraper, bot, web spider, atau web crawler. Teknik-teknik yang umumnya digunakan untuk web scraping antara lain, Parsing HTML, Parsing DOM, XPath, dan Google Docs (Web Scraping: Panduan dan Teknik-Tekniknya, 2020). Instagram crawler adalah salah satu teknik web scraping yang dikhususkan untuk mengambil data pada platform media sosial Instagram. Konsep dari Instagram crawler juga cukup sederhana, yaitu mengambil data seperti Instagram post, profile, hashtag tanpa menggunakan Instagram API [6].

### 2.5 Instagram Fake Spammer Genuine Accounts Dataset

Dataset yang digunakan didapatkan dari kaggle.com, yang bernama “*Instagram Fake Spammer Genuine Accounts*” [4]. Isi dari *dataset* ini merupakan kumpulan akun *Instagram* yang terdiri dari 12 atribut yang telah diproses oleh pembuat. Atribut-atribut tersebut adalah gambar profil, rasio angka pada *username*, rasio angka pada *fullname*, jumlah kata pada *fullname*, sama atau tidaknya *username* dan *fullname*, panjang deskripsi, ada atau tidaknya *external url*, jumlah *post*, *following*, *followers*, dan label *fake* atau *genuine*.

### 2.6 Waikato Environment for Knowledge Analysis

Waikato Environment for Knowledge Analysis atau yang dikenal sebagai WEKA adalah sebuah perangkat lunak yang dapat menerapkan berbagai algoritma machine learning untuk melakukan proses data mining. Fitur-fitur yang umumnya dimanfaatkan pada WEKA di antaranya adalah *classification*, *regression*, *clustering*, *association rules*, *visualization*, dan *data preprocessing*. Fitur pada WEKA yang akan dimanfaatkan untuk penelitian ini adalah *data preprocessing* untuk membagi dataset pada data training dan data testing [14].

### 2.7 Pengujian Confusion Matrix

Pengujian *confusion matrix* adalah sebuah metode pengujian yang digunakan untuk mengukur kinerja suatu metode klasifikasi binary. Cara kerja *confusion matrix* adalah dengan membandingkan hasil klasifikasi yang dilakukan oleh algoritma klasifikasi dengan data yang sebenarnya [13]. Ada empat istilah untuk hasil klasifikasi, diantaranya:

Tabel 1. Empat Istilah untuk Hasil Klasifikasi

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	(TP) <i>True Positive</i>	(FN) <i>False Negative</i>
Negatif	(FP) <i>False Positive</i>	(TN) <i>True Negative</i>

Hasil dari klasifikasi yang merupakan *decision rules* akan dilakukan pengujian dengan menggunakan *confusion matrix* untuk mengetahui tingkat akurasi dan sensitivitasnya. Akurasi adalah

persentase jumlah data yang diprediksi secara benar terhadap jumlah seluruh data [11]. Perhitungan tingkat akurasi akan dihitung menggunakan persamaan 3 dan perhitungan sensitifitas akan diperhitungkan menggunakan persamaan 4.

$$Accuracy = \frac{tp+tn}{tp+fp+fn+tn} \quad (3)$$

$$Recall (Sensitivity) = \frac{tp}{tp+fn} \quad (4)$$

Dimana  $tp$  merupakan jumlah kasus *true positive*,  $tn$  merupakan kasus *true negative*,  $fp$  merupakan kasus *false positive*, dan  $fn$  merupakan kasus *false negative* pada data.

### 3. DESAIN SISTEM

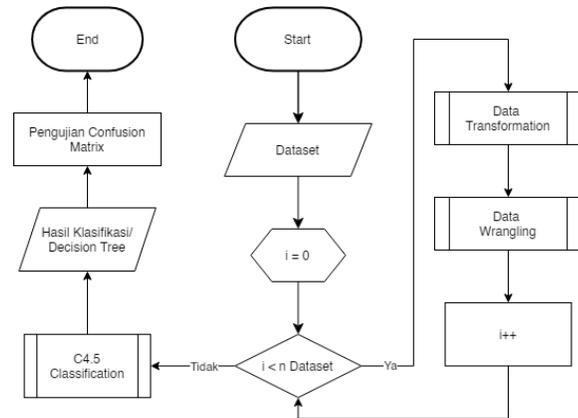
#### 3.1 Analisis Permasalahan

Penelitian tentang prediksi kredibilitas akun dilakukan karena seiring berjalannya waktu, *platform* media sosial *Instagram* semakin berkembang. Dibalik berkembangnya *Instagram*, semakin banyak juga pihak yang membuat akun dengan tujuan mencari keuntungan pribadi dengan merugikan orang lain. Untuk menciptakan rasa aman dan nyaman saat berjejaring sosial di media sosial *Instagram*, ditelitilah sebuah sistem untuk melakukan prediksi kredibilitas akun *Instagram*.

#### 3.2 Desain Sistem

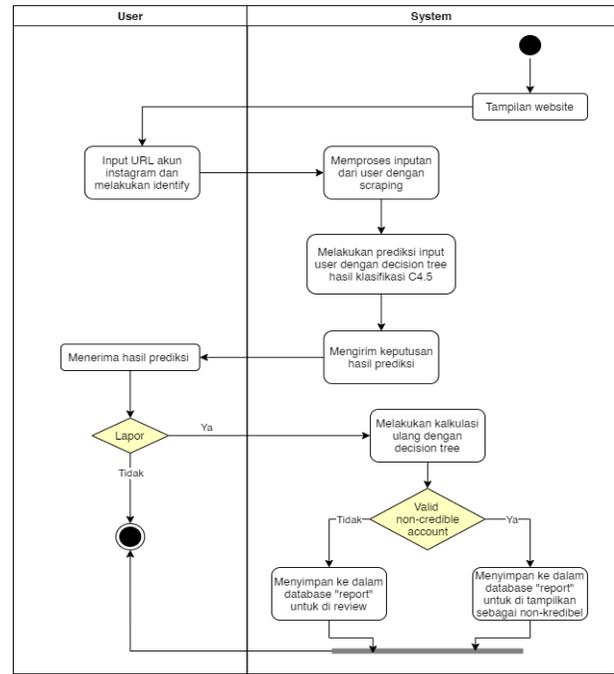
Pembuatan web identifikasi terdiri dari dua proses utama yaitu *data classification* dan *data testing*. *Data classification* akan menggunakan metode klasifikasi C4.5. *Data testing* merupakan proses pengujian prediksi oleh kumpulan data dengan hasil klasifikasi pada *data classification* yang menghasilkan *decision rules*.

Langkah pertama, file dataset yang berformat *CSV* akan dibaca kedalam *IDE Python* dengan menggunakan *library Pandas*. Obyek data yang masuk ke dalam *IDE* akan terbaca sebagai format *dataframe*. Obyek inilah yang akan diproses dengan klasifikasi C4.5 dengan menggunakan *library Chefboost*. Hasil terbaik berdasarkan *data testing* akan digunakan untuk pembuatan sistem web.



Gambar 1. Flowchart Data Training

Gambaran website yang akan dibangun adalah website dengan kumpulan informasi tentang sebuah akun yang akan diinputkan oleh *user*. *User* melakukan input suatu akun *Instagram*, lalu sistem akan memproses dan mencari kredibilitas akun yang telah *user input*, lalu melakukan umpan balik kredibilitas yang dihasilkan. Gambar 1 adalah proses penggunaan website yang digambarkan melalui *activity diagram*.



Gambar 2. Activity Identifikasi Akun

#### 3.2.1 Data Training

Data yang akan digunakan untuk data *training* dan *testing* adalah sebanyak 576 (lima ratus tujuh puluh enam data). Data tersebut didapat dari sumber pada *kaggle.com* yang bernama “*Instagram Fake Spammer Genuine Accounts*”, namun hanya disebutkan bahwa data-data dalam dataset didapat dari *proses machine learning* yang telah dibuat oleh pembuatnya. Kumpulan data ini telah memiliki label berupa *fake account* atau *genuine account*.

#### 3.2.2 Data Preparation

Data preparation dilakukan pada dataset karena atribut-atribut yang terdapat masih memiliki nilai yang beragam dan inkonsisten. Hal ini dapat membuat proses klasifikasi menjadi kurang maksimal, maka data preparation dilakukan untuk menghasilkan nilai akurasi yang lebih baik. Proses ini terbagi menjadi dua bagian yaitu, data transformation dan data wrangling.

##### 3.2.2.1 Data Transformation

Merupakan proses dimana perubahan terhadap data dengan format tertentu dilakukan. Proses ini bertujuan agar data tersebut dapat lebih sesuai untuk tahap klasifikasi. Terdapat dua jenis transformasi yang akan dilakukan yaitu transformasi nilai dan transformasi bentuk. Transformasi nilai merupakan proses penambahan sebuah atribut baru pada dataset. Transformasi bentuk adalah proses merubah bentuk dari data atribut yang berupa numerik menjadi bentuk diskrit [2]. Berikut adalah perincian dari kedua transformasi:

1. Transformasi nilai dengan menambahkan atribut rasio antara *Followers* dan *Following*. Perhitungan rasio ini dapat diamati pada persamaan 5.
$$Rasio\ Followers\ per\ Following = \frac{Jumlah\ Followers}{Jumlah\ Following} \quad (5)$$
2. Transformasi bentuk pada atribut *Followers*, *Following*, dan *Rasio Followers per Following*. Persamaan 6 hingga persamaan 14 adalah persamaan yang membahas mengenai transformasi bentuk pada tiap atribut yang ada.
$$Followers = Low, Jumlah\ Followers < 300 \quad (6)$$

$$Followers = High, Jumlah\ Followers \geq 300 \quad (7)$$

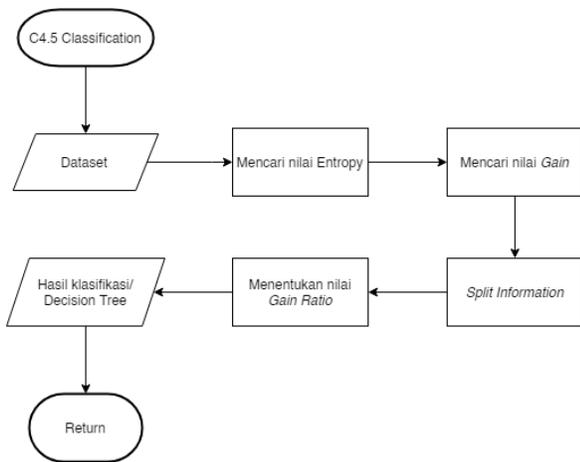
- Followers = Low, Jumlah Following < 500 (8)
- Following = High, Jumlah Following  $\geq$  500 (9)
- Ratio = Spammer, Ratio < 0.5 (10)
- Ratio = Suspicious,  $0.5 \leq$  Ratio < 1 (11)
- Ratio = Normal,  $1 \leq$  Ratio < 2 (12)
- Ratio = Micro,  $2 \leq$  Ratio < 10 (13)
- Ratio = Influencer, Ratio > 10 (14)

### 3.2.2.2 Data Wrangling

Pada Proses *data wrangling* dataset yang akan digunakan untuk *training* dan *testing* akan diimpor menggunakan *library Pandas* pada *Python IDE*. Selain proses impor, yaitu memasukan file berformat *.csv*, *library Pandas* juga berperan dalam

### 3.3 C4.5 Classification

*Data classification* dengan metode klasifikasi C4.5 akan diterapkan kepada model yang telah melalui proses *data preparation*. Proses klasifikasi akan dilakukan menggunakan bantuan *Chefboost library* pada bahasa pemrograman *Python*. *Library Chefboost* akan mempelajari karakteristik dari model dan akan membangun sebuah *decision tree* hasil klasifikasi.



Gambar 3. Flowchart Klasifikasi C4.5

### 3.4 Desain Pengujian Klasifikasi

Pengujian akan dilakukan dengan 4 kelompok data dengan jumlah antara *data training* dan *data testing* yang berbeda [3], dan masing-masing terdiri dari 3 buah model dengan data yang berbeda. Jumlah data yang ada sebanyak 576 (lima ratus tujuh puluh enam). Data tersebut terdiri dari 288 (dua ratus delapan puluh delapan) data positif, yaitu data dengan label akun *fake* atau *spammer* dan 288 (dua ratus delapan puluh delapan) data negatif, yaitu data dengan label akun yang *genuine* menurut dataset. Kelompok data tersebut antara lain:

1. Kelompok A, yaitu 90% *data training* :10% *data testing*
2. Kelompok B, yaitu 80% *data training* : 20% *data testing*
3. Kelompok C, yaitu 75% *data training* : 25% *data testing*
4. Kelompok D, yaitu 67% *data training* : 33% *data testing*

Langkah-langkah pengujian yang dilakukan antara lain:

1. Mempersiapkan masing-masing kelompok data untuk dilakukan training. Data ini diambil dari kumpulan dataset sesuai pembagian yang telah ditentukan.

2. Melakukan implementasi training dengan *library Chefboost* dan mencari decision rules dari masing-masing kelompok data dan model.
3. Melakukan prediksi kepada masing-masing *data testing*, dan melakukan pencatatan terhadap variabel-variabel yang diperlukan oleh *confusion matrix* (*true positive*, *true negative*, *false positive*, dan *false negative*).
4. Melakukan perhitungan akurasi dan sensitivitas tiap model menggunakan persamaan *confusion matrix*, lalu menyimpulkan akurasi dan sensitivitas tertinggi.

## 4. PENGUJIAN SISTEM

Pengujian klasifikasi C4.5 dalam sistem ini dilakukan bertahap, dimulai dari kelompok A dengan jumlah *data training* 518 data dan *data testing* 58 data. Kelompok B dengan jumlah *data training* 460 data dan *data testing* 116 data. Kelompok C dengan jumlah *data training* 432 data dan *data testing* 144 data. Kelompok D dengan jumlah *data training* 384 data dan *data testing* 192 data.

### 4.1 Hasil Pengujian

Berdasarkan pengujian yang telah dilakukan pada masing masing kelompok, hasilnya dapat dijabarkan sebagai berikut:

Tabel 2. Hasil Pengujian Kelompok Uji A

Klasifikasi Data Training Kelompok A			
Model	Time	Accuracy	Recall
a	7.5178	0.7931	0.8621
b	6.3035	0.8966	0.8966
c	7.0944	0.8966	0.931
Mean	6.9719	0.8621	0.896566667

Berdasarkan Tabel 2, pengujian memerlukan rata-rata waktu 6,9 detik untuk melakukan klasifikasi dengan akurasi rata-rata 0,8621 dan sensitivitas 0,8966.

Tabel 3. Hasil Pengujian Kelompok Uji B

Klasifikasi Data Training Kelompok B			
Model	Time	Accuracy	Recall
a	8.4284	0.9707	0.8793
b	7.8436	0.9052	0.8966
c	6.1932	0.9224	0.8621
Mean	7.4884	0.932766667	0.879333333

Berdasarkan Tabel 3, pengujian memerlukan rata-rata waktu 7,4 detik untuk melakukan klasifikasi dengan akurasi rata-rata 0,9327 dan sensitivitas 0,8793.

Tabel 4. Hasil Pengujian Kelompok Uji C

Klasifikasi Data Training Kelompok C			
Model	Time	Accuracy	Recall
a	6.3379	0.8819	0.8611
b	5.6885	0.8889	0.8472
c	5.2735	0.9028	0.8889
Mean	5.766633333	0.8912	0.865733333

Berdasarkan Tabel 4, pengujian memerlukan rata-rata waktu 5,76 detik untuk melakukan klasifikasi dengan akurasi rata-rata 0,8912 dan sensitivitas 0,8657.

**Tabel 5. Hasil Pengujian Kelompok Uji D**

Klasifikasi Data Training Kelompok D			
Model	Time	Accuracy	Recall
a	6.6364	0.8906	0.8438
b	4.3237	0.8698	0.9167
c	5.374	0.8802	0.8333
Mean	5.4447	0.8802	0.8646

Berdasarkan Tabel 5, pengujian memerlukan rata-rata waktu 5,44 detik untuk melakukan klasifikasi dengan akurasi rata-rata 0,8802 dan sensitifitas 0,8646.

## 4.2 Decision Rules

Berdasarkan nilai akurasi dan sensitivitas terbaik yang didapatkan dari pengujian klasifikasi, maka ditentukanlah *decision rules* yang akan dijadikan acuan dalam pembuatan web identifikasi. *Decision rules* kelompok uji B model a, karena merupakan model dengan hasil akurasi tertinggi dan sensitivitas diatas rata-rata. *Decision rules* yang dihasilkan memiliki *root* atribut *profile picture* dilanjutkan dengan panjang deskripsi pada profil dan pengaruh angka pada *username* pada cabang pertama dan kedua.

## 5. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan pengujian dapat disimpulkan bahwa, keakuratan metode klasifikasi C4.5 juga bergantung kepada jumlah data training yang digunakan. Pada penelitian ini akurasi dan nilai sensitivitas terbaik ditemukan pada kumpulan perbandingan data training dan data testing 80:20 dengan total jumlah 576 data. Akurasi yang didapatkan sebesar 97,07% dan nilai sensitivitas atau recall sebesar 87,93%.

Decision tree hasil klasifikasi C4.5 yang telah dilakukan, telah ditemukan beberapa faktor yang cukup mempengaruhi suatu kredibilitas sebuah akun Instagram. Faktor yang paling mempengaruhi di antaranya yaitu ada atau tidaknya *profile picture* pada akun, disusul dengan jumlah karakter pada *description* dan pengaruh angka pada *username* akun *Instagram*.

### 5.2 Saran

Saran yang dapat digunakan untuk pengembangan aplikasi web, maupun penelitian serupa untuk lebih lanjut diantaranya:

1. Dalam detail kredibilitas bagian poin-poin apa saja yang dapat ditingkatkan agar akun menjadi lebih kredibel, akan lebih baik apabila user mendapat informasi lebih detail mengenai solusinya, seperti menyebutkan jumlah *followers* yang harus ditingkatkan.
2. Pelaporan terhadap suatu akun tertentu akan memiliki bobot lapor yang berbeda bagi tiap pengguna yang melapor. Dalam contoh kasus seorang pengguna yang akun Instagram-nya memiliki indikasi akun yang sangat kredibel, laporannya akan lebih bisa diterima jika dibandingkan dengan pengguna yang memiliki akun Instagram yang baru saja dibuat.
3. Sistem pelaporan timbal balik antara akun yang diidentifikasi dinilai perlu. Apabila ada akun yang teridentifikasi sebagai akun yang tidak kredibel, namun seharusnya adalah kredibel, maka pengguna dapat melaporkan bahwa akun tersebut kredibel.

## 6. DAFTAR PUSTAKA

- [1] Ahmed, S. 2016. Evaluating Presto as an SQL on Hadoop solution: A Case at Truecaller. URI= <http://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-47369>
- [2] Anggitasari, P. 2016. Klasifikasi Spammer Pada Twitter Berdasarkan Perilaku Pengguna Menggunakan Algoritma C5.0. (Jan. 2016), 6-14. URI= <http://repository.ipb.ac.id/handle/123456789/79820>
- [3] Arifman, P. 2020. Klasifikasi Bot Pada Media Sosial Twitter Menggunakan Algoritma C4.5. URI= <http://repository.uin-suska.ac.id/id/eprint/30651>
- [4] Bakhshandeh, B. 2020. Instagram Fake Spammer Genuine Accounts. URI= <https://www.kaggle.com/free4ever1/instagram-fake-spammer-genuine-accounts>
- [5] Bronshtein, A. 2017. A Quick Introduction to the "Pandas" Python Library. URI= <https://towardsdatascience.com/a-quick-introduction-to-the-pandas-python-library-f1b678f34673>
- [6] Erin. 2020. (Jan. 2020), Instagram Followers: How Many Does the Average Person Have? Retrieved from Hastags For Likes. URI= <https://www.hashtagsforlikes.co/blog/instagram-followers-how-many-does-the-average-person-have/>
- [7] Huaying - Instagram Crawler. (n.d.). URI= <https://github.com/huaying/instagram-crawler>
- [8] Irena, B. and Erwin Budi Setiawan 2020. Fake News (Hoax) Identification on Social Media Twitter using Decision Tree C4.5 Method. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informatika)*. 4, 4 (Aug. 2020), 711 - 716. DOI= <https://doi.org/10.29207/resti.v4i4.2125>
- [9] Kamagi, D. and Hansun, S. 2014. Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Ultimatics : Jurnal Teknik Informatika*. 6, 1 (Jun. 2014), 15-20. DOI= <https://doi.org/https://doi.org/10.31937/ti.v6i1.327>.
- [10] Kurniawan, Y. I. 2018. Perbandingan Algoritma Naive Bayes dan C.45 Dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 5, 4 (May. 2018), 457-458. DOI= <http://dx.doi.org/10.25126/jtiik.201854803>
- [11] Lestari, D. P. 2015. Analisis Strategi Internet Marketing Butik Online di Surabaya Melalui Instagram. *Commonline*, 4, 2 (Jan. 2015), 412-424.
- [12] Rahayu, W. and Wahyudi, E. 2017. Classical Test Theory of Innapropriate Index Score's Accuracy Comparison Using Confusion Matrix Accuracy Proportion in Educational Measurement. *Indonesian Journal of Educational Review*, 4, 1 (Jul. 2017), 84-92. DOI= <https://doi.org/10.21009/IJER.04.01.08>
- [13] Serengil, S. I. 2018. Chefboost. URI= <https://github.com/serengil/chefboost>
- [14] Suhartono, D. 2018. Weka: Software untuk Memahami Konsep Data Mining. Retrieved from Binus University School of Computer Science. (Nov. 2018), URI= <https://socs.binus.ac.id/2018/11/29/weka-software-untuk-memahami-konsep-data-mining/>