

Prediksi Kepribadian Pengguna Instagram Berdasarkan Model Big Five Personality Menggunakan Algoritma SVM

Muhamad Ramadhan¹, Ricky Eka Putra²

^{1,2}Jurusan Teknik Informatika/Teknik Informatika, Universitas Negeri Surabaya

¹muhamadramadhan16051204017@mhs.unesa.ac.id

²rickyeka@unesa.ac.id

Abstrak— Popularitas media sosial Instagram terus merangkak naik sejak debutnya pada tahun 2010. Popularitas tersebut membuat Instagram menjadi media sosial terbesar di dunia di samping Facebook dan Twitter. Kini sebagian besar perusahaan mulai menggunakan media sosial sebagai salah satu aspek untuk menilai kepribadian pelamar. Analisis kepribadian dapat dilakukan dengan mengamati aktivitas di media sosial. Hal ini menjadi alasan dibalik pentingnya pembuatan sistem prediksi kepribadian otomatis berdasarkan aktivitas di media sosial. Instagram dipilih sebagai sumber data pada penelitian ini, dan *Big Five Personality* dipilih sebagai model kepribadian untuk pelabelan data. Pengumpulan data dilakukan dengan melakukan tes kepribadian terhadap 191 pengguna Instagram, kemudian melakukan *crawling feed* akun Instagram menggunakan API Instagram. Berdasarkan hasil pengujian sistem, kombinasi algoritma *Support Vector Machine* dengan kernel RBF dan *Bayesian Optimization* mampu menghasilkan akurasi mencapai 60.34%, presisi 30.17%, dan *recall* 50% untuk mengklasifikasi kepribadian pengguna Instagram.

Kata Kunci— prediksi kepribadian, *big five personality*, instagram, *support vector machine*, *bayesian optimization*.

I. PENDAHULUAN

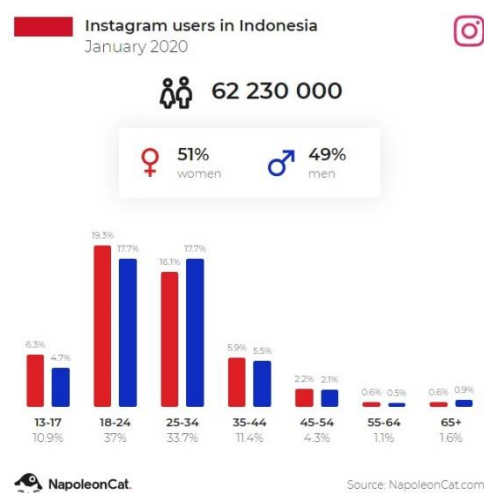
Popularitas media sosial Instagram terus merangkak naik sejak debutnya pada tahun 2010. Bahkan Instagram masuk ke dalam 7 media sosial terbesar versi situs *Search Engine Journal* [1]. Berdasarkan pantauan terakhir dari situs NapoleonCat pada Januari 2020, jumlah pengguna aktif bulanan Instagram di Indonesia mencapai 62.230.000 orang. Jumlah tersebut merupakan 22,8% dari total populasi penduduk di Indonesia.

Gbr 1 menunjukkan bahwa pengguna aktif bulanan Instagram di Indonesia didominasi oleh perempuan dengan persentase sebesar 51%. Alasan dibalik itu telah diungkap oleh Alice Marwick, seorang peneliti media sosial di Universitas Fordham, bahwa Instagram dapat menarik perhatian perempuan karena fokus pada konten visual, yaitu gambar dan video [2].

Kini sebagian besar perusahaan mulai menggunakan media sosial sebagai salah satu aspek untuk menilai kepribadian pelamar. Analisis kepribadian dilakukan dengan mengamati aktivitas di media sosial. Mengetahui kepribadian pelamar adalah hal yang penting bagi perusahaan untuk menentukan penempatan kerja yang tepat bagi pelamar tersebut. Apabila pekerjaan tidak sesuai dengan kepribadian, maka perusahaan akan mengalami kerugian.

Ada banyak model kepribadian yang dapat digunakan untuk memprediksi kepribadian, seperti DISC (*Dominance Influence Steadiness Conscientiousness*), MBTI (*Myers-Briggs Type*

Indicator), dan *Big Five Personality* [3]. Di antara ketiga model tersebut, *Big Five Personality* menjadi model kepribadian terpopuler dan paling banyak digunakan di dunia psikologi [4].



Gbr 1. Statistik Pengguna Aktif Instagram di Indonesia [5]

Big Five Personality pertama kali muncul pada tahun 1960 oleh Allport dan Cattell [6]. Model *Big Five Personality* membagi kepribadian ke dalam 5 kategori kepribadian, yaitu *Openness to Experience* (terbuka dengan hal-hal baru), *Conscientiousness* (kehati-hatian), *Extroversion* (ekstroversi), *Agreeableness* (keramahan), dan *Neuroticism* (neurotisme). Salah satu instrumen tes untuk mengukur kepribadian seseorang dengan model *Big Five Personality* yaitu *International Personality Item Pool – Big Five Factor Markers 50* (IPIP-BFM-50) yang berisi 50 pertanyaan [4]. Tabel 1 menjelaskan karakteristik dari *Big Five Personality* [7].

Tabel 1
Karakteristik *Big Five Personality*

Kepribadian	Rendah	Tinggi
<i>Openness</i> (O)	Konvensional, minat sempit, sederhana, tertutup, konservatif, tidak artistik.	Modern, minat luas, kreatif, penasaran, original, imajinatif.
<i>Conscientiousness</i> (C)	Lalai, malas, tidak dapat diandalkan, berkemauan lemah, ceroboh, hedonistik.	Teratur, disiplin, rapi, tepat waktu, pekerja keras, ambisius, dapat diandalkan, teliti.
<i>Extroversion</i> (E)	Menyendiri, pasif, mengutamakan tugas.	Bersosialisasi, aktif, penyayang.

Kepribadian	Rendah	Tinggi
	pendiam, tenang, tidak bersemangat, pendiam.	mengutamakan orang lain, optimis, banyak bicara, menyenangkan.
Agreeableness (A)	Kasar, kejam, curiga, tidak kooperatif, dendam, mudah tersinggung, manipulatif, sinis.	Lemah lembut, baik, kooperatif, pemaaf, alami, membantu, mudah tertipu, lugas.
Neuroticism (N)	Tidak emosional, tenang, santai, kuat secara emosional, <i>secure</i> , puas terhadap dirinya sendiri.	Emosional, khawatir, <i>insecure</i> , canggung, gugup, hipokondria

Penelitian terkait prediksi kepribadian dengan model *Big Five Personality* melalui media sosial pernah dilakukan oleh Agnes dan Masayu [8]. Penelitian tersebut menggunakan 119 pengguna Twitter usia kisaran 19–25 tahun yang masing-masing diambil 200 tweet terbarunya sebagai dataset. Data yang telah dikumpulkan menggunakan API Twitter akan diolah untuk memperoleh fitur-fitur yang menggambarkan perilaku sosial pengguna Twitter. Sedangkan untuk pelabelan data, dilakukan melalui tes kepribadian menggunakan kuesioner. Dengan algoritma *Support Vector Regression*, diperoleh nilai *Mean Absolute Error* (MAE) terkecil 0,2739.

Penelitian lain untuk melakukan prediksi kepribadian melalui media sosial juga pernah dilakukan oleh Muhammad Fikry dan Yusra [9]. Dataset yang digunakan adalah 46 akun pengguna Twitter. Dengan algoritma *Support Vector Machine*, akurasi yang diperoleh mencapai 88,89%.

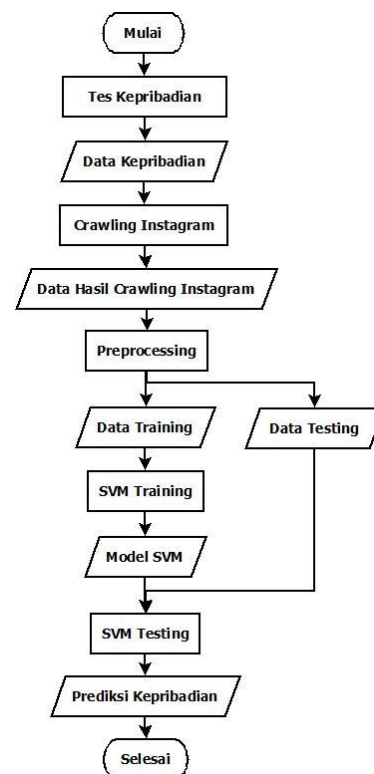
Kedua penelitian terdahulu yang menggunakan *Support Vector Machine* (SVM) sebagai algoritma klasifikasi dapat memberikan hasil akurasi yang cukup tinggi [10]. Cara kerja dari SVM yaitu dengan menemukan fungsi keputusan (*hyperplane*) terbaik yang dapat memisahkan antar kelas data. Dengan diketahuinya *hyperplane* yang paling optimal, SVM dapat mengklasifikasi data yang belum diketahui kelasnya dengan melihat data tersebut berada di sisi mana dari *hyperplane*.

Berdasarkan uraian diatas, maka penelitian ini akan berfokus pada implementasi algoritma SVM untuk melakukan prediksi kepribadian pengguna Instagram berdasarkan model *Big Five Personality*. Pengumpulan data dilakukan dengan melakukan tes kepribadian terhadap pengguna Instagram, kemudian melakukan *crawling feed* akun Instagram sebagai fitur data menggunakan API Instagram. Sedangkan SVM dipilih sebagai algoritma untuk melakukan klasifikasi data.

II. METODOLOGI

Penelitian ini bertujuan untuk menerapkan algoritma *Support Vector Machine* (SVM) pada sistem prediksi kepribadian pengguna Instagram. Alur kerja dari penelitian ini yaitu melakukan tes kepribadian terhadap responden yang memiliki akun Instagram, kemudian mengambil data *feed* Instagram responden melalui proses *crawling* dan *preprocessing* sebagai ekstraksi fitur untuk melatih SVM, lalu langkah terakhir adalah melatih SVM agar SVM dapat

melakukan prediksi. Untuk meningkatkan kinerja SVM, ditambahkan algoritma *Bayesian Optimization* seperti [11]. Penelitian ini mengadopsi model *Big Five Personality* untuk memberi label kepribadian pada pengguna Instagram. Alur kerja dari penelitian ini dapat dilihat pada Gbr 2.



Gbr 2. Alur Kerja Penelitian

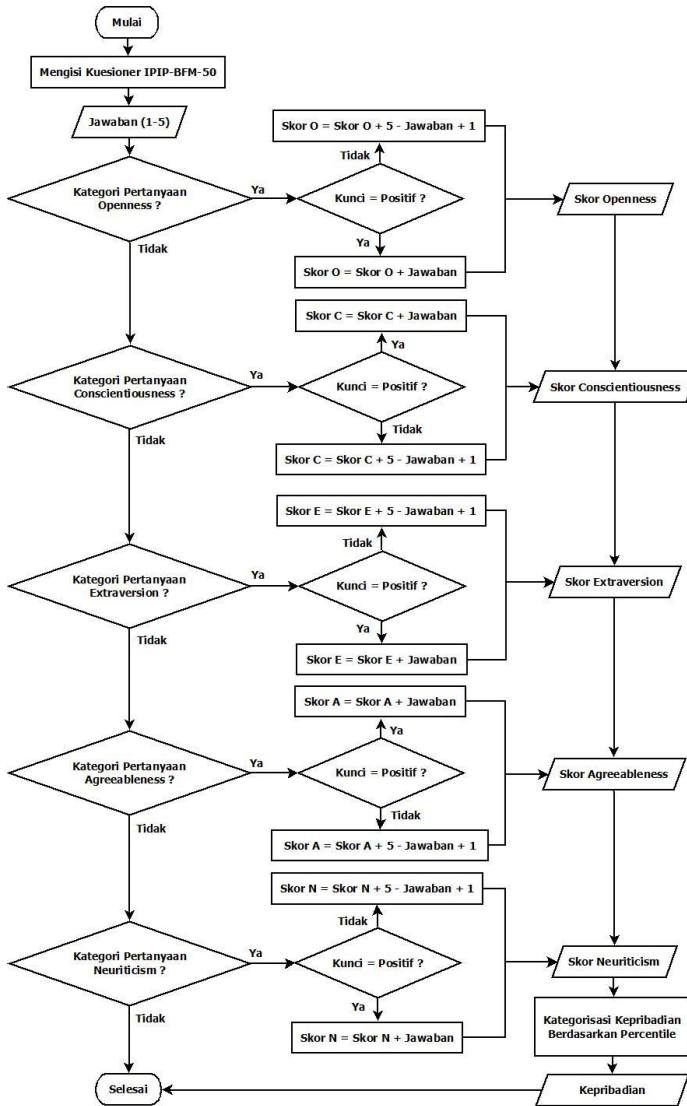
A. Tes Kepribadian

Untuk mengetahui kepribadian dari pengguna Instagram, dilakukan tes kepribadian menggunakan instrumen IPIP-BFM-50 yang telah diterjemahkan ke dalam bahasa Indonesia, berisi 50 pertanyaan dengan jawaban berupa nilai skala 1-5. Jawaban dari 50 pertanyaan pada IPIP-BFM-50 dihitung dengan teknik perhitungan skor sesuai panduan di situs resmi IPIP beralamat https://ipip.ori.org/New_IPIP-50-item-scale.htm.

Tabel 2
Contoh Butir Pertanyaan pada Kuesioner IPIP-BFM-50

Pertanyaan	Kategori	Kunci
Saya menghidupkan suasana dalam suatu acara	E	+
Saya tidak terlalu memedulikan orang lain	A	-
Saya selalu mempersiapkan segala hal	C	+
Saya mudah merasa tertekan	N	-
Saya menguasai banyak kosakata	O	+
Saya tidak banyak berbicara	E	-
Saya peduli dengan orang lain	A	+
Saya meninggalkan barang pribadi di sembarang tempat	C	-
Saya merasa tenang hampir setiap saat	N	+
Saya kesulitan memahami ide yang bersifat abstrak	O	-

IPIP-BFM (*International Personality Item Pool – Big Five Factor Marker*) adalah salah satu alat ukur *open-source* untuk mengukur *Big Five Personality* yang banyak digunakan oleh para peneliti di dunia [4]. IPIP-BFM pernah digunakan di dalam penelitian Goldberg [12]. IPIP-BFM memiliki 2 versi, yaitu IPIP-BFM yang terdiri dari 100 butir pertanyaan (IPIP-BFM-100), dan IPIP-BFM yang terdiri dari 50 butir pertanyaan (IPIP-BFM-50).



Gbr 3. Alur Kerja Tes Kepribadian

Masing-masing butir pertanyaan pada IPIP-BFM-50 memiliki kategori, yaitu O untuk *Openness to Experience*, C untuk *Conscientiousness*, E untuk *Extroversion*, A untuk *Agreeableness*, dan N untuk *Neuroticism*. Tiap butir pertanyaan juga memiliki kunci skor "+" dan "-" seperti pada Tabel 2. Apabila pertanyaan dengan tipe O dan kunci "+" dijawab dengan nilai 5, maka skor kepribadian *Openness to Experience* akan bertambah 5. Sedangkan pertanyaan dengan tipe O dan kunci "-" dijawab dengan nilai 5, maka skor kepribadian *Openness to Experience* hanya bertambah 1, karena kunci "-"

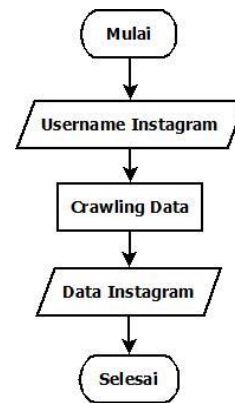
mengindikasikan bahwa perhitungan skor dilakukan secara terbalik.

Raw score atau skor mentah dari hasil tes kepribadian masing-masing kategori kepribadian akan dikategorisasi menjadi 2 kelas, yaitu tinggi dan rendah. Kategorisasi tersebut dilakukan dengan mengonversi *raw score* menjadi *z-score* (istilah umum dalam statistika untuk menyatakan besaran jarak suatu nilai dari sampel terhadap *mean* dalam satuan standar deviasinya), kemudian menghitung nilai distribusi normalnya [13]. Nilai distribusi normal di bawah 50% akan dikategorikan menjadi rendah, sedangkan nilai distribusi normal di atas sama dengan 50% akan dikategorikan menjadi tinggi.

Alur tes kepribadian dijelaskan secara rinci pada Gbr 3. Pada tahap ini, *input* yang masuk adalah jawaban dari tes kepribadian terhadap pengguna Instagram, sedangkan *output* yang keluar adalah label/kelas kepribadian.

B. Crawling Data Instagram

Setelah label kepribadian didapatkan, dilakukan *crawling* data menggunakan API Instagram terhadap akun pengguna Instagram yang telah melakukan tes kepribadian. *Crawling* adalah proses mengumpulkan atau mengunduh data dari suatu *database* [14]. Data hasil *crawling* lalu dilabeli dengan hasil tes kepribadian, sehingga diperoleh dataset untuk digunakan di dalam penelitian ini.



Gbr 4. Alur Kerja *Crawling* Instagram

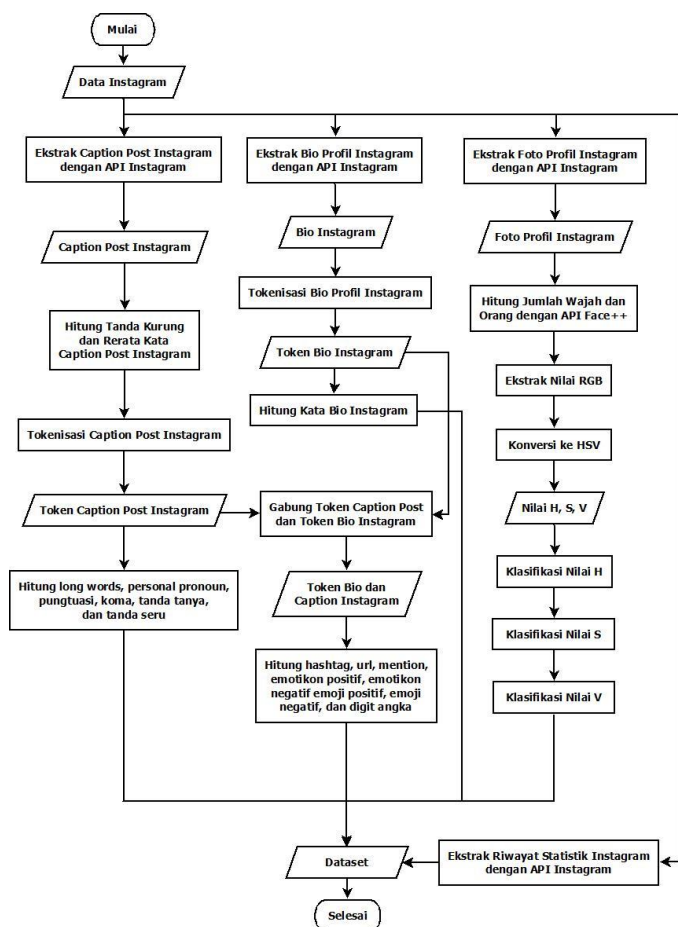
Alur *crawling* Instagram dijelaskan secara rinci pada Gbr 4. Pada tahap ini, *input* yang masuk adalah *username* Instagram, sedangkan *output* yang keluar adalah data instagram seperti pada Tabel 3.

Tabel 3
Output dari *Crawling* Instagram

No.	Data Hasil <i>Crawling</i> Instagram yang Digunakan
1.	Foto profil
2.	Deskripsi atau bio profil
3.	URL eksternal pada bio profil
4.	Jumlah <i>follower</i>
5.	Jumlah <i>following</i>
6.	Jumlah postingan
7.	Jumlah komentar
8.	Jumlah <i>likes</i>
9.	Jumlah <i>video views</i>
10.	Isi <i>caption</i> postingan

C. Preprocessing Data

Data hasil *crawling* merupakan dataset mentah yang belum siap digunakan, sehingga data tersebut memerlukan *preprocessing* untuk memperoleh fitur-fitur lebih lengkap yang dibutuhkan di dalam penelitian ini. Tahap *preprocessing* dijabarkan secara rinci pada Gbr 5.



Gbr 5. Alur Kerja Preprocessing

Tahap *preprocessing* terdiri dari beberapa proses, seperti tokenisasi, penerapan regular expression (regex), penerapan API Face++ untuk mengekstraksi jumlah wajah (*face*) dan orang (*full-body*) pada foto profil, serta melakukan konversi citra RGB ke citra HSV pada foto profil untuk mengekstraksi fitur visual HSV. *Output* dari tahap *preprocessing* dapat dilihat pada Tabel 4. Sebagian dari fitur pada Tabel 4 mengacu pada [8], [15], [16], [17], [18] dan [19].

Tabel 4
Output dari Preprocessing

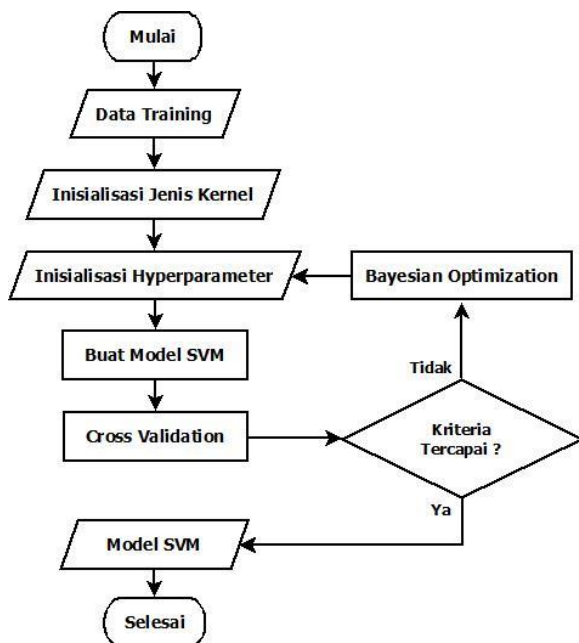
No.	Fitur	Keterangan
1.	Jumlah <i>follower</i>	Jumlah orang yang mengikuti
2.	Jumlah <i>following</i>	Jumlah orang yang diikuti
3.	Jumlah <i>likes</i>	Jumlah orang yang menyukai postingan
4.	Jumlah komentar	Jumlah komentar pada seluruh postingan

No.	Fitur	Keterangan
5.	Jumlah <i>video views</i>	Jumlah orang yang menonton postingan berupa video
6.	Jumlah gambar	Jumlah postingan berupa gambar
7.	Jumlah video	Jumlah postingan berupa video
8.	Jumlah <i>sidecar</i>	Jumlah postingan yang bisa di- <i>swipe</i>
9.	Jumlah kata pada bio	Jumlah kata pada deskripsi/bio profil
10.	Jumlah <i>hashtag</i>	Jumlah kata di dalam <i>caption</i> post dan bio yang mengandung <i>hashtag</i> atau jumlah kata dengan panjang lebih dari 2 huruf yang diawali tanda "#"
11.	Jumlah url	Jumlah eksternal url pada bio ditambah dengan jumlah kata berawalan "http://" dan "https://" pada <i>caption</i> post
12.	Jumlah <i>mention</i>	Jumlah pengguna lain yang di <i>mention</i> di dalam <i>caption</i> post atau jumlah kata dengan panjang lebih dari 2 huruf yang diawali dengan tanda @
13.	Jumlah emotikon positif	Jumlah emotikon positif pada <i>caption</i> post dan bio
14.	Jumlah emotikon negatif	Jumlah emotikon negatif pada <i>caption</i> post dan bio
15.	Jumlah emoji positif	Jumlah emoji positif pada <i>caption</i> post dan bio
16.	Jumlah emoji negatif	Jumlah emoji negatif pada <i>caption</i> post dan bio
17.	Jumlah angka	Jumlah digit angka pada <i>caption</i> post dan bio
18.	Rerata kata tiap post	Rata-rata dari jumlah kata pada <i>caption</i> post
19.	<i>Long words</i>	Jumlah kata dengan panjang lebih dari 6 huruf pada <i>caption</i> post
20.	<i>Personal Pronoun</i>	Jumlah kata yang berisi kata "saya", "aku", "kamu", "kau", "kami", "mereka", "dia" pada <i>caption</i> post
21.	Jumlah <i>punctuation</i>	Jumlah kata yang mengandung <i>punctuation</i> pada <i>caption</i> post
22.	Jumlah koma	Jumlah koma atau "," pada <i>caption</i> post
23.	Jumlah tanda seru	Jumlah tanda seru atau "!" pada <i>caption</i> post
24.	Jumlah tanda tanya	Jumlah tanda Tanya atau "?" pada <i>caption</i> post
25.	Jumlah tanda kurung	Jumlah tanda kurung atau "(" pada <i>caption</i> post
26.	Jumlah wajah	Jumlah wajah yang terdeteksi pada foto profil
27.	Jumlah orang	Jumlah orang <i>full-body</i> yang terdeteksi pada foto profil
28.	<i>Hue</i>	Nilai <i>hue</i> yang diperoleh dari foto profil kemudian digolongkan ke dalam kategori warna <i>warm</i> (0-80° dan 330-360°) atau warna <i>cold</i> (81-329°)
29.	<i>Saturation</i>	Nilai <i>saturation</i> yang diperoleh dari foto profil kemudian digolongkan ke dalam kategori <i>low</i> (0-30%), <i>mid</i> (31-69%), atau <i>high</i> (70-100%)
30.	<i>Value</i>	Nilai <i>value</i> yang diperoleh dari foto profil kemudian digolongkan ke dalam kategori

No.	Fitur	Keterangan
		low (0-30%), mid (31-69%), atau high (70-100%)

D. Training Data

SVM akan melakukan proses *training* terhadap data *training*. Data *training* diperoleh dari dataset yang telah dibagi menjadi data *training* dan data *testing*. *Bayesian Optimization* juga berperan di dalam proses *training* untuk melakukan pencarian *hyperparameter* terbaik dari SVM. Alur proses *training* SVM dijabarkan seperti pada Gbr 6.



Gbr 6. Alur Kerja Proses *Training* SVM dengan *Bayesian Optimization*

Sebelum melakukan *training* menggunakan SVM, jenis kernel dan nilai *hyperparameter* diinisialisasi terlebih dahulu. Kernel yang umum digunakan pada SVM adalah kernel Linier, *Sigmoid*, *Polynomial*, dan RBF [20]. *Hyperparameter* yang umum diinisialisasi pada SVM adalah nilai *cost* dan nilai *gamma*. Nilai *cost* pada SVM memengaruhi *margin* dan *error* dalam proses klasifikasi, sedangkan nilai *gamma* pada SVM merupakan *hyperparameter* dari kernel RBF untuk mengatasi permasalahan non-linier.

Pemilihan pasangan *hyperparameter cost* dan *gamma* yang tepat dapat meningkatkan akurasi dari SVM. Menguji kombinasi *hyperparameter* SVM satu per satu sampai menemukan akurasi tertinggi akan memakan banyak waktu dan tenaga. Untuk mengatasi masalah ini, dibutuhkan algoritma untuk mencari *hyperparameter* terbaik seperti *Grid Search* atau *Bayesian Optimization*.

Dibandingkan dengan *Grid Search*, *Bayesian Optimization* memiliki waktu komputasi yang relatif lebih singkat dalam melakukan pencarian *hyperparameter*. Hal ini karena *Grid Search* bekerja dengan menguji semua kombinasi *hyperparameter* yang memungkinkan dari yang telah

ditentukan, sedangkan *Bayesian Optimization* hanya menguji beberapa kombinasi *hyperparameter* dalam ruang interval yang telah ditentukan. *Bayesian Optimization* dapat dipilih sebagai algoritma pencarian *hyperparameter* apabila ruang interval *hyperparameter* yang ingin dicari terlalu besar.

Setelah menemukan *hyperparameter* terbaik, proses *training* SVM akan menghasilkan model SVM yang paling optimal untuk mengklasifikasi data pada proses *testing*. Model SVM dapat ditulis dalam bentuk persamaan (1).

$$f(\phi(x)) = \text{sign}(w \cdot \phi(x) + b) \quad (1)$$

Dimana :

$f(\phi(x))$ = fungsi keputusan

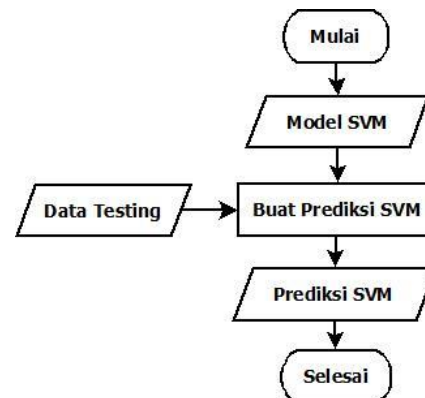
w = bobot

$\phi(x)$ = fungsi kernel

b = bias

E. Testing Data

Model SVM yang dihasilkan dari proses *training* akan digunakan untuk menghasilkan prediksi terhadap data *testing* saat proses *testing* SVM seperti Gbr 7. Dengan model SVM, SVM dapat memprediksi kepribadian pengguna Instagram.



Gbr 7. Alur Kerja Proses *Testing* SVM

F. Evaluasi Performa Sistem

Uji performansi sistem dilakukan dengan melakukan perhitungan nilai akurasi, presisi, dan *recall*. Nilai akurasi digunakan untuk mengukur tingkat kedekatan antara nilai aktual dengan nilai prediksi, nilai presisi digunakan untuk mengukur tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan oleh sistem, sedangkan nilai *recall* digunakan untuk mengukur tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

Perhitungan nilai akurasi dilakukan dengan rumus seperti persamaan (2).

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Dimana :

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

Perhitungan nilai presisi dilakukan dengan rumus seperti persamaan (3).

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (3)$$

Sedangkan perhitungan nilai *recall* dilakukan dengan rumus seperti persamaan (4).

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

Dalam praktiknya, algoritma SVM akan dilatih untuk menghasilkan model yang dapat memprediksi kepribadian dari pengguna Instagram, kemudian saat melakukan prediksi akan menghasilkan nilai *True Positive* (data *true* yang diprediksi *true*), *False Negative* (data *true* yang diprediksi *false*), *False Positive* (data *false* yang diprediksi *true*), dan *True Negative* (data *false* yang diprediksi *false*). Sebagai contoh, apabila model SVM menghasilkan *True Positive* 8, *False Negative* 7, *False Positive* 3, dan *True Negative* 12, maka akurasi dapat dihitung menggunakan rumus pada persamaan (2). Contoh perhitungan akurasi dapat dilihat pada persamaan (5).

$$\text{Akurasi} = \frac{8 + 12}{8 + 7 + 3 + 12} = \frac{20}{30} = 0.67 \text{ (67\%)} \quad (5)$$

III. HASIL PEMBAHASAN

Data yang digunakan di dalam penelitian ini adalah data *feed* instagram yang diperoleh dari pengguna instagram yang telah melakukan tes kepribadian. Sistem prediksi kepribadian yang dibangun di dalam penelitian ini akan memprediksi tinggi rendahnya tiap kategori kepribadian *Big Five Personality* secara terpisah, yaitu *Openness to Experience* (O), *Conscientiousness* (C), *Extroversion* (E), *Agreeableness* (A), dan *Neuroticism* (N). Data instagram diperoleh melalui proses *crawling* menggunakan API Instagram, sedangkan kelas atau label data diperoleh melalui tes kepribadian. Aplikasi yang dibangun pada penelitian ini menggunakan *framework* PHP Laravel dan *database* PostgreSQL.

A. Tes Kepribadian

Instrumen kuesioner yang digunakan untuk melakukan tes kepribadian dengan model kepribadian *Big Five Personality* adalah IPIP-BFM-50. Kuesioner dibuat di Google Form dan dibagikan secara umum melalui Instagram, Whatsapp, dan LINE. Di dalam kuesioner juga ditambahkan kolom isian untuk menulis *username* instagram dari responden. Setelah target jumlah responden tercapai, hasil tanggapan kuesioner di Google Form akan diunduh dengan format CSV.

Hasil kepribadian responden diperoleh dengan melakukan perhitungan skor kepribadian sesuai aturan IPIP-BFM-50 sebagaimana ditunjukkan pada Gbr 3. Aplikasi yang dibangun pada penelitian ini memiliki fitur untuk melakukan perhitungan skor kepribadian secara otomatis sesuai alur pada Gbr 3. Fitur

tersebut hanya membutuhkan *import file* CSV yang diperoleh dari Google Form untuk bekerja. Hal tersebut dilakukan untuk memangkas waktu karena jumlah responden yang banyak.

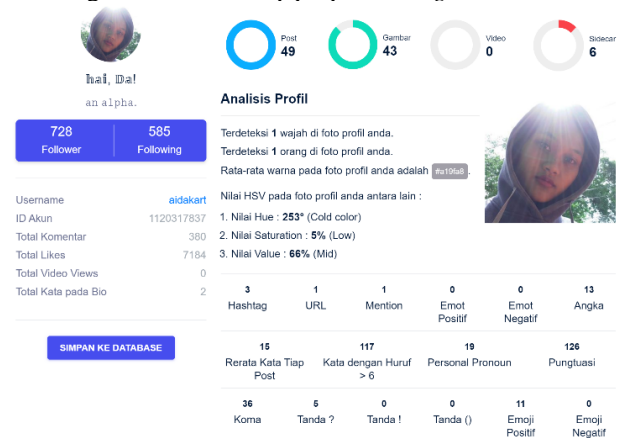
Setelah melakukan proses perhitungan skor kepribadian, diperoleh data hasil kepribadian seperti pada Tabel 5. Setiap responden memiliki 5 data kepribadian yang merepresentasikan tinggi rendahnya masing-masing kategori *Big Five Personality*.

Tabel 5
Data Hasil Tes Kepribadian

	Tinggi	Rendah	Total
<i>Openness to Experience</i>	88	103	191
<i>Conscientiousness</i>	99	92	191
<i>Extroversion</i>	97	94	191
<i>Agreeableness</i>	88	103	191
<i>Neuroticism</i>	94	97	191

B. Crawling Data Instagram

Data *feed* instagram dari responden yang telah melakukan tes kepribadian akan diambil melalui proses *crawling* menggunakan API Instagram. Proses *crawling* membutuhkan *username* instagram yang telah ditulis oleh responden saat mengisi kuesioner sebagai *input*. Hasil *crawling* instagram akan menghasilkan data seperti pada Tabel 3 untuk tiap pengguna instagram. Data hasil *crawling* kemudian dikembangkan melalui tahap *preprocessing*.



Gbr 8. Contoh Hasil *Crawling* Setelah Melalui Tahap *Preprocessing*

Crawling data terhadap 191 pengguna instagram dilakukan pada tanggal 30-31 Mei tahun 2020. Lamanya proses *crawling* tergantung pada beberapa faktor, seperti jumlah *post* yang diunggah dan jumlah *caption* setiap *post*. Contoh hasil *crawling* dapat dilihat pada Gbr 8, dimana data jumlah *follower*, *following*, dan fitur *feed* instagram lainnya tertera dalam format angka. Data-data tersebut akan disimpan ke dalam *database*.

C. Preprocessing Data Instagram

Preprocessing data instagram berfokus pada ekstraksi fitur dari foto profil, biografi, dan *caption* dari pengguna instagram. Dalam implementasinya, *preprocessing* dilakukan tepat setelah *crawling* sehingga menjadi satu kesatuan proses. Fitur yang dapat diekstraksi dari foto profil adalah fitur visual dan fitur konten. Sedangkan fitur yang dapat diekstraksi dari *caption* dan

biografi yaitu fitur perilaku sosial seperti jumlah *mention*, URL, *hashtag*, emotikon, emoji, *pungtuasi*, *personal pronoun*, *long word*, angka, tanda koma, tanda kurung, tanda tanya, dan tanda seru.

Fitur visual maupun konten dari sebuah gambar sama-sama baik dan dapat digunakan untuk memprediksi kepribadian [16]. Aplikasi yang dibangun pada penelitian ini memanfaatkan API Face++ untuk mendeteksi jumlah wajah dan orang pada foto profil sebagai fitur konten. P. Sedangkan nilai HSV dari foto profil merupakan fitur visual. Ekstraksi nilai RGB dilakukan terlebih dahulu pada foto profil, kemudian nilai RGB akan dikonversi menjadi HSV. Masing-masing nilai H, S, dan V akan diklasifikasikan menjadi beberapa kategori berdasarkan interval nilai seperti pada Tabel 6.

Tabel 6
Klasifikasi Nilai HSV

Parameter	Interval Nilai	Kategori
H (Hue)	$0^\circ \leq H \leq 80^\circ$	Warm Color (1)
	$330^\circ \leq H \leq 360^\circ$	Warm Color (1)
	$81^\circ \leq H \leq 329^\circ$	Cold Color (0)
S (Saturation)	$0\% \leq S \leq 30\%$	Low (0)
	$31\% \leq S \leq 69\%$	Middle (1)
	$70\% \leq S \leq 100\%$	High (2)
V (Value)	$0\% \leq V \leq 30\%$	Low (0)
	$31\% \leq V \leq 69\%$	Middle (1)
	$70\% \leq V \leq 100\%$	High (2)

Ekstraksi fitur pada *caption* dan biografi pengguna instagram dilakukan dengan cara melakukan tokenisasi, *case folding*, dan *Regular Expression* (Regex). Beberapa fitur dapat diekstraksi hanya dengan perintah *if else* tanpa menggunakan Regex, seperti *mention* yang dapat dideteksi dengan memeriksa keberadaan simbol @ di awal *token*. Semua fitur yang diperoleh dari proses *crawling* dan *preprocessing* akan disimpan ke dalam *database*.

D. Implementasi Support Vector Machine

SVM diimplementasikan menggunakan *library* PHP RubixML. SVM digunakan untuk melakukan proses *training* dan *testing* terhadap dataset. Sedangkan dataset akan dibagi menjadi 70% data *training* dan 30% data *testing* secara acak dan *10-Fold Cross Validation*.

Penelitian ini menggunakan jenis kernel RBF untuk proses *training* SVM. Dengan penggunaan jenis kernel RBF, proses *training* SVM juga membutuhkan beberapa *hyperparameter* seperti nilai *cost* dan *gamma*. Nilai *cost*, *gamma*, dan jenis kernel akan disimpan sebagai sebagai model SVM setelah proses *training* dilakukan.

Tabel 7
Hasil Pencarian *Hyperparameter* SVM menggunakan *Bayesian Optimization* untuk kepribadian *Openness to Experience*

Iterasi ke-	Akurasi	Cost	Gamma
1	43,63%	0.4448	0.1517
15	46,59%	1.007	9.215
29	45%	7.816	8.942
84	52,03%	1.147	8.036
99	46,48%	3.479	6.008

<i>Hyperparameter</i> yang terpilih	1.147	8.036
-------------------------------------	-------	-------

Tabel 8
Hasil Pencarian *Hyperparameter* SVM menggunakan *Bayesian Optimization* untuk kepribadian *Conscientiousness*

Iterasi ke-	Akurasi	Cost	Gamma
3	35,22%	2.584	2.655
37	41,21%	7.69	6.851
59	38,35%	9.956	0.104
84	46,59%	8.026	4.461
100	39,89%	1.226	1.966
<i>Hyperparameter</i> yang terpilih	8.026	4.461	

Tabel 9
Hasil Pencarian *Hyperparameter* SVM menggunakan *Bayesian Optimization* untuk kepribadian *Extroversion*

Iterasi ke-	Akurasi	Cost	Gamma
2	5,11%	6.497	0.1653
31	41,87%	9.689	1.525
63	45,27%	7.601	0.1614
78	39,89%	3.446	0.1188
90	51,43%	5.645	0.1138
<i>Hyperparameter</i> yang terpilih	5.645	0.1138	

Tabel 10
Hasil Pencarian *Hyperparameter* SVM menggunakan *Bayesian Optimization* untuk kepribadian *Agreeableness*

Iterasi ke-	Akurasi	Cost	Gamma
4	39,18%	8.979	2.914
18	51,1%	3.83	4.524
43	41,26%	1.381	6.507
66	41,54%	10	1.326
97	42,14%	9.991	2.378
<i>Hyperparameter</i> yang terpilih	3.83	4.524	

Tabel 11
Hasil Pencarian *Hyperparameter* SVM menggunakan *Bayesian Optimization* untuk kepribadian *Neuroticism*

Iterasi ke-	Akurasi	Cost	Gamma
8	39,18%	5.5	6.43
24	46,76%	6.542	0.4864
46	43,63%	6.638	8.554
75	40,55%	6.138	0.1
92	46,7%	1.555	7.826
<i>Hyperparameter</i> yang terpilih	6.542	0.4864	

Penelitian ini menggunakan *Bayesian Optimization* sebagai algoritma untuk mencari *hyperparameter* SVM terbaik. *Bayesian Optimization* digunakan untuk mencari fungsi objektif maksimal berdasarkan *hyperparameter* yang dapat diperoleh. *Bayesian Optimization* menciptakan suatu model pengganti dari fungsi objektif menggunakan metode regresi. Model pengganti yang umum digunakan adalah *Gaussian Process*. Lalu fungsi akuisi digunakan sebagai penyeleksi titik sampel untuk dievaluasi berikutnya. Fungsi akuisi yang digunakan pada penelitian ini adalah *Upper Confidence Bound* (UCB). Implementasi *Bayesian Optimization* pada penelitian ini menggunakan *library* Python *FMFN Bayesian Optimization*.

Tabel 7, Tabel 8, Tabel 9, Tabel 10, dan Tabel 11 merupakan gambaran dari hasil pencarian *hyperparameter* SVM terbaik pada penelitian ini. Pencarian tersebut dilakukan menggunakan

Bayesian Optimization dengan 100 iterasi. Hyperparameter terbaik yang dicari pada penelitian ini yaitu nilai *cost* pada interval 0,01 sampai 10 dan nilai *gamma* pada interval 0,01 sampai 10. Berdasarkan pengujian yang telah dilakukan, Bayesian Optimization dapat mengevaluasi performa setiap pasangan *cost* dan *gamma* dari SVM berdasarkan interval yang telah ditentukan untuk menemukan akurasi tertinggi.

Nilai *cost* dan *gamma* yang diperoleh dari Bayesian Optimization akan digunakan untuk menginisialisasi hyperparameter pada proses *training* untuk menciptakan model *hyperplane* dari SVM. Selanjutnya SVM akan melakukan prediksi terhadap data *testing*. Perhitungan akurasi, presisi, dan *recall* dilakukan untuk mengetahui performa yang dihasilkan oleh algoritma SVM pada penelitian ini. Performa SVM pada penelitian ini dapat dilihat pada Tabel 12.

Tabel 12
Hasil Akurasi, Presisi, dan Recall

Kepribadian	Akurasi	Presisi	Recall
<i>Openness to Experience</i>	58.62%	29.31%	50%
<i>Conscientiousness</i>	53.45%	26.72%	50%
<i>Extroversion</i>	50%	25%	50%
<i>Agreeableness</i>	60.34%	30.17%	50%
<i>Neuroticism</i>	51.72%	25.86%	50%

Berdasarkan percobaan yang telah dilakukan, diperoleh hasil akurasi 58.62%, presisi 29.31%, dan *recall* 50% untuk kepribadian *Openness to Experience*; akurasi 53.45%, presisi 26.72%, dan *recall* 50% untuk kepribadian *Conscientiousness*; akurasi 50%, presisi 25%, dan *recall* 50% untuk kepribadian *Extroversion*; akurasi 60.34%, presisi 30.17%, dan *recall* 50% untuk kepribadian *Agreeableness*; serta akurasi 51.72%, presisi 25.86%, dan *recall* 50% untuk kepribadian *Neuroticism*.

Penelitian ini mengungkapkan peran kombinasi hyperparameter SVM dalam melakukan prediksi kepribadian pengguna Instagram. Kombinasi hyperparameter yang tepat dapat meningkatkan akurasi dari hasil prediksi SVM. Semakin tinggi akurasi yang dihasilkan oleh model SVM, maka proses prediksi kepribadian pengguna Instagram akan semakin baik. Percobaan menggunakan dataset yang berbeda akan menghasilkan hasil akurasi, presisi, dan *recall* yang berbeda, sehingga penelitian lain yang serupa dapat memperoleh hasil yang berbeda dari penelitian ini.

IV. KESIMPULAN

Berdasarkan percobaan yang telah dilakukan di dalam penelitian ini, didapatkan kesimpulan bahwa :

1. Penerapan model *Big Five Personality* terhadap prediksi kepribadian pengguna Instagram dapat dilakukan dengan melakukan tes kepribadian menggunakan instrumen kuesioner IPIP-BFM-50.
2. Sistem prediksi kepribadian pengguna Instagram berdasarkan model *Big Five Personality* menggunakan algoritma *Support Vector Machine* dapat diimplementasikan menjadi sebuah aplikasi dengan bahasa pemrograman PHP, *database* PostgreSQL, API Instagram,

API Face++, library PHP RubixML, dan library Python *FMFN Bayesian Optimization*.

3. Sistem prediksi kepribadian pengguna Instagram berdasarkan model *Big Five Personality* menggunakan algoritma *Support Vector Machine* pada penelitian ini menghasilkan akurasi 58.62%, presisi 29.31%, dan *recall* 50% untuk kepribadian *Openness to Experience*; akurasi 53.45%, presisi 26.72%, dan *recall* 50% untuk kepribadian *Conscientiousness*; akurasi 50%, presisi 25%, dan *recall* 50% untuk kepribadian *Extroversion*; akurasi 60.34%, presisi 30.17%, dan *recall* 50% untuk kepribadian *Agreeableness*; serta akurasi 51.72%, presisi 25.86%, dan *recall* 50% untuk kepribadian *Neuroticism*. Kombinasi hyperparameter yang tepat dapat meningkatkan akurasi dari hasil prediksi SVM. Semakin tinggi akurasi yang dihasilkan oleh model SVM, maka proses prediksi kepribadian pengguna Instagram akan semakin baik.

REFERENSI

- [1] K. Kellogg, "The 7 Biggest Social Media Sites in 2020," 3 Februari 2020. [Online]. Available: <https://www.searchenginejournal.com/>. [Använd 12 Februari 2020].
- [2] H. Seligson, "Why Are More Women Than Men on Instagram?," 7 Juni 2016. [Online]. Available: <https://www.theatlantic.com/technology/archive/2016/06/why-are-more-women-than-men-on-instagram/485993/>.
- [3] T. Tandra, Hendro, D. Suhartono, R. Wongso and Y. L. Prasetyo, "Personality Prediction System from Facebook Users," *2nd International Conference on Computer Science and Computational Intelligence*, 2017.
- [4] H. Akhtar and S. Azwar, "Indonesian Adaptation and Psychometric Properties Evaluation of the Big Five Personality Inventory: IPIP-BFM-50," *Psikologi*, vol. 46, nr 1, p. 32–44, 2019.
- [5] NapoleonCat, "Instagram Users in Indonesia," Januari 2020. [Online]. Available: <https://napoleoncat.com/stats/instagram-users-in-indonesia/2020/01>. [Använd 13 Februari 2020].
- [6] Maulanski, "Kepribadian Big Five: Variabel Psikologi," 9 Desember 2016. [Online]. Available: <https://psikologihore.com/kepribadian-big-five/>.
- [7] L. A. Pervin, D. Cervone and O. P. John, *Personality: Theory and research*, USA: John Wiley & Sons, 2005.
- [8] A. T. Damanik and M. L. Khodra, "Prediksi Kepribadian Big 5 Pengguna Twitter dengan Support Vector Regression," *Cybermatika*, vol. 3, nr 1, pp. 14-22, 2015.
- [9] M. Fikry and Yusra, "Ekstrover atau Introversi : Klasifikasi Kepribadian Pengguna Twitter dengan Menggunakan Metode Support Vector Machine," *Sains, Teknologi dan Industri*, pp. 72-76, 2018.
- [10] B. Fahmi, A. Wibowo and D. Hajar, "Analisa Kepribadian Pengguna Facebook Menggunakan Algoritma Support Vector Machine," *Komputer Terapan*, vol. 5, nr 1, p. 28–35, 2019.
- [11] G. Y. N. Adi, M. H. Tandio, V. Ong and D. Suhartono, "Optimization for Automatic Personality Recognition," *3rd International Conference on Computer Science and Computational Intelligence*, pp. 473-480, 2018.
- [12] L. R. Goldberg, "The Development of Markers for the Big-Five Factor Structure.," *Psychological Assessment*, vol. 4, nr 1, pp. 26-42, 1992.
- [13] S. Srivastava, "Norms for the Big Five Inventory and other personality measures," *The Hardest Science*, 17 Oktober 2012. [Online]. Available: <https://thehardestscience.com/2012/10/17/norms-for-the-big-five-inventory-and-other-personality-measures/>. [Använd 2 Juni 2020].

- [14] J. E. Sembodo, E. B. Setiawan och Z. A. Baizal, "Data Crawling Otomatis pada Twitter," i *Indonesia Symposium on Computing*, Bandung, 2016.
- [15] C. Segalin, F. Celli, L. Polonio, M. Kosinski, D. Stillwell, N. Sebe, M. Cristani och B. Lepri, "What your Facebook Profile Picture Reveals about your Personality," *25th ACM International Conference on Multimedia*, pp. 460-468, 2017.
- [16] B. Ferwerda och M. Tkalcic, "Predicting Users' Personality from Instagram Pictures: Using Visual and/or Content Features?," *UMAP '18: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, pp. 157-161, 2018.
- [17] B. Ferwerda, M. Schedl och M. Tkalcic, "Using Instagram Picture Features to Predict Users' Personality," *Lecture Notes in Computer Science*, vol. 9516, pp. 850-861, 2016.
- [18] F. Celli, "Mining User Personality in Twitter," *Language, Interaction and Computation Laboratory (CLIC)*, 2011.
- [19] J. Golbeck, C. Robles, M. Edmondson och K. Turner, "Predicting Personality from Twitter," *IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing*, pp. 149-156, 2011.
- [20] Hefinioanrhys, "Support Vector Machines with the MLR Package," 10 Oktober 2019. [Online]. Available: <https://www.r-bloggers.com/support-vector-machines-with-the-mlr-package/>.