

Sistem Prediksi Kepribadian Big Five Personality Berdasarkan Data Pengguna Facebook

Tommy Tandra

Computer Science Department
School of Computer Science
Bina Nusantara University
Indonesia

tommy.tandra@binus.ac.id

Hendro

Computer Science Department
School of Computer Science
Bina Nusantara University
Indonesia

hendro004@binus.ac.id

Derwin Suhartono

Computer Science Department
School of Computer Science
Bina Nusantara University
Indonesia

dsuhartono@binus.edu

Rini Wongso

Computer Science Department
School of Computer Science
Bina Nusantara University
Indonesia

rwongso@binus.edu

Yen Lina Prasetyo

Computer Science Department
School of Computer Science
Bina Nusantara University
Indonesia

yenlina@binus.edu

Abstract

Penggunaan jaringan sosial saat ini telah mencapai puncaknya. Berbagai informasi dibagikan secara luas melalui sosial media seperti Facebook. Informasi mengenai pengguna dan status yang dibagikan merupakan suatu aset penting untuk dijadikan bahan penelitian di bidang pembelajaran perilaku dan kepribadian manusia. Berbagai penelitian sejenis telah dilakukan di bidang ini dan terus berkembang hingga saat ini. Penelitian ini mencoba untuk membangun sebuah sistem yang dapat memprediksi kepribadian seseorang berdasarkan informasi pengguna Facebook. Penelitian ini menerapkan beberapa teknik machine learning dan juga mencoba teknik deep learning untuk kemudian dilakukan analisa komprehensif metode yang memiliki akurasi paling baik. Hasil yang ditunjukkan pada penelitian ini berhasil mengungguli hasil akurasi penelitian sejenis sebelumnya dengan akurasi tertinggi 93.33% menggunakan arsitektur deep learning dan 79.33% untuk teknik machine learning.

1 Pendahuluan

Social media telah menjadi sebuah alat komunikasi dan interaksi antar manusia selama beberapa tahun terakhir. Di zaman dimana hampir seluruh manusia memiliki *smartphone*, interaksi langsung antar manusia sudah jarang terjadi. Sehingga sulit untuk mengenali dan mengetahui kepribadian dari seseorang. Tetapi, hal ini berbeda dengan apa yang terjadi di *social media*. Facebook memiliki pengguna terbesar

yang mencapai 1,8 Milyar pengguna dengan sekitar 800 juta user menghabiskan sekitar 40 menit sehari menggunakan aplikasi Facebook ini (Bachrach et al., 2012). Pengguna Facebook pada umumnya mengungkapkan ekspresi, perasaan maupun opini mereka di *user feed* mereka. Walaupun saat ini Facebook lebih banyak digunakan untuk berbagi foto dan video, penelitian kali ini akan berfokus pada *linguistic* pengguna.

Berbagai penelitian di bidang psikologi menunjukkan bahwa terdapat korelasi antara kepribadian dengan tingkah laku *linguistic* (Ilmu bahasa) seseorang. Korelasi ini dapat dengan efektif dianalisa dan digambarkan dengan pendekatan NLP (*Natural Language Processing*). Oleh sebab itu penelitian ini dilakukan untuk membuat sebuah sistem prediksi kepribadian seseorang melalui data aktivitas pengguna tersebut di *social media* Facebook. Sistem prediksi ini akan dibangun dengan menggunakan model kepribadian "*Big Five Personality Traits*". Terdapat beberapa model kepribadian lain seperti MBTI (*Myers-Briggs Type Indicator*) atau DISC. Tetapi, *Big Five Personality Traits* merupakan model yang sedang populer saat ini dan digunakan dalam banyak penelitian yang berhubungan dengan kepribadian seseorang. *Traits* yang ada dalam Big Five ini meliputi *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, dan *Neuroticism*.

Penelitian ini akan menggunakan dataset sebanyak 250 *user* dari myPersonality (Kosinski et al., 2015) dan dataset yang didapatkan secara manual sebanyak 150 *user*. Sistem prediksi yang

akan dibangun akan menggunakan teknik algoritma classifier machine learning untuk proses training. Penelitian sejenis yang menggunakan teknik machine learning sebagai classifier telah banyak dilakukan, namun penggunaan deep learning pada bidang ini masih cukup sulit ditemukan. Oleh karena itu, penelitian ini juga akan melakukan implementasi teknik deep learning sebagai model classifier untuk melihat hasil dari pengguna teknik ini. Implementasi teknik yang menghasilkan nilai akurasi terbaik selama percobaan akan dijadikan sebagai model untuk sistem prediksi kepribadian penelitian ini.

2 Penelitian Terkait

Penelitian sebelumnya mengenai prediksi kepribadian telah dilakukan oleh Farnadi et al. (2013) menggunakan sosial media Facebook dan beberapa fitur seperti LIWC, fitur *Social Network*, fitur *Time-related*, dan fitur lainnya. Schwartz et al. (2013) melakukan penelitian mengenai prediksi kepribadian berdasarkan status Facebook dengan menggunakan dua pendekatan yaitu *open-vocabulary* DLA (*Differential Language Analysis*) dan fitur LIWC.

Alam et al. (2013) melakukan penelitian menggunakan sosial media Facebook dan pendekatan *bag-of-words* dan menggunakan token (*unigrams*) sebagai fiturnya. Penelitian lainnya yang dilakukan oleh Wijaya et al. (2016) membuat sebuah sistem prediksi kepribadian menggunakan sosial media Twitter dengan LIWC dan MRC sebagai fiturnya.

Fernadi et al. (2013), Schwartz et al. (2013), Alam et al. (2013), dan Wijaya et al. (2016) melakukan penelitian mengenai prediksi kepribadian dengan menggunakan sosial media dalam bahasa Inggris dan model kepribadian *Big Five Personality Traits*. Penelitian baru-baru ini yang dilakukan oleh Ong et al. (2017) membuat sebuah sistem prediksi kepribadian berdasarkan model kepribadian *Big Five Personality Traits* dengan menggunakan sosial media Twitter dalam bahasa Indonesia.

Penelitian lain mengenai prediksi kepribadian juga pernah dilakukan oleh Majumder et al. (2013) yang menggunakan teknik *Deep Learning* untuk mengklasifikasikan model kepribadian *Big Five Personality Traits* dengan menggunakan sosial media Facebook.

3 Metodologi

3.1 Dataset

Dataset yang digunakan pada penelitian ini dibagi menjadi 2. Dataset pertama didapatkan dari myPersonality (Kosinski et al., 2015) sebanyak 250 dataset user Facebook dengan kurang lebih 10.000 status yang telah diberikan pelabelan kepribadian berdasarkan model kepribadian *Big Five Personality Traits*. Distribusi dataset myPersonality berdasarkan jenis kepribadiannya dipresentasikan pada Tabel 1 di bawah.

Value	OPN	CON	EXT	AGR	NEU
Yes	176	130	96	134	99
No	74	120	154	116	151

Tabel 1 Distribusi dataset myPersonality

Dataset kedua adalah postingan status dari 150 dataset user Facebook yang didapatkan dengan pengumpulan manual oleh peneliti yang kemudian disebut dataset *Manual Gathering*. Graph API Facebook dimanfaatkan dalam proses pengumpulan dataset tersebut. Pelabelan kepribadian kemudian dilakukan dengan memasukkan postingan user secara manual ke dalam aplikasi *applymagicsauce*. Tabel 2 merupakan hasil distribusi dataset setelah diberi pelabelan kepribadian berdasarkan model kepribadian *Big Five Personality Traits*.

Value	OPN	CON	EXT	AGR	NEU
Yes	97	63	38	81	50
No	53	87	112	69	100

Tabel 2 Distribusi dataset *Manual Gathering*

3.2 Fitur-fitur yang digunakan

Penelitian ini menggunakan beberapa fitur untuk melihat perbandingan hasil dan kemampuannya. Fitur yang digunakan dibedakan untuk masing-masing implementasi pembelajaran. Untuk implementasi *machine learning* akan menggunakan fitur linguistik dengan pendekatan *closed-vocabulary*. *Closed vocabulary* merupakan fitur yang berbasis dengan jumlah kandungan kata-kata sesuai dengan fitur yang telah ditentukan (*predefined features*). Untuk pendekatan ini, peneliti menggunakan fitur

linguistik LIWC (Pennebaker, 2015) dan SPLICE (Moffit et al. 2012).

LIWC yang digunakan pada penelitian ini merupakan versi LIWC2015 yang memiliki 85 fitur yang telah dikembangkan dari versi LIWC2007. Pada penelitian ini fitur LIWC yang akan digunakan sebanyak 85 fitur.

SPLICE merupakan fitur linguistik yang diciptakan oleh Moffit et al. dan telah digunakan di beberapa penelitian di bidang ini. Pada penelitian ini fitur SPLICE yang akan digunakan sebanyak 74 fitur.

Selain kedua fitur linguistik di atas, penelitian ini juga akan memanfaatkan penggunaan fitur *Social Network Analysis* yang diberikan oleh dataset myPersonality yang berupa informasi detail mengenai jaringan pertemanan seorang pengguna. Untuk informasi lengkap mengenai fitur ini dapat dilihat pada (O'Malley & Marsden, 2008).

Berbeda dengan implementasi *machine learning*, implementasi *deep learning* akan dilakukan terpisah dengan menggunakan fitur linguistik pendekatan *open vocabulary*. *Open vocabulary* tidak membutuhkan fitur yang telah ditentukan sebelumnya (*predefined features*). Pendekatan ini akan melakukan eksplorasi otomatis terhadap dataset yang digunakan untuk mencari hubungan antar penggunaan kata dengan kepribadian. Penelitian sebelumnya yang melakukan perbandingan antara kedua pendekatan fitur linguistik ini adalah (Schwartz et al., 2013).

3.3 Preprocessing

Data yang telah dikumpulkan dalam penelitian ini akan melalui tahapan preprocessing terlebih dahulu sebelum dilakukan pengujian untuk membangun model klasifikasi. Tahap *preprocessing* yang dilakukan yaitu *remove URLs*, *remove symbols*, *remove names*, *remove spaces*, *lower case text*, *stemming*, and *remove stopwords*.

Khusus untuk status dengan bahasa Indonesia, dilakukan proses *preprocessing* tambahan secara manual dengan mengganti penggunaan *slang words* atau kata-kata tidak baku dari status terlebih dahulu untuk kemudian dilanjutkan ke tahap penerjemahan ke dalam bahasa Inggris.

3.4 Klasifikasi Model

Implementasi *machine learning* menggunakan 5 algoritma *classifier* berbeda yaitu Naive Bayes, *Support Vector Machine* (SVM), *Logistic Regression*, *Gradient Boosting*, dan *Linear Discriminant Analysis* (LDA). Untuk validasi model, peneliti menggunakan teknik *10-fold cross validation* dengan menggunakan library Python scikit-learn. *10-fold cross validation* membagi 10% dataset menjadi data *testing* dan 90% dataset sebagai data *training* secara bergantian.

Peneliti melakukan serangkaian uji coba dengan berbagai skenario untuk melihat bagaimana hasil akurasi setiap algoritma tersebut dalam melakukan prediksi terhadap jenis kepribadian. Uji coba dilakukan dengan menambahkan beberapa proses tambahan untuk meningkatkan akurasi. Proses pertama yang dilakukan adalah dengan *Features Selection* yang mencoba untuk melakukan filterisasi atau penghapusan fitur-fitur yang digunakan yang dianggap memiliki korelasi rendah terhadap *traits* dari kepribadian. Proses yang dilakukan selanjutnya adalah melakukan proses *Resampling* yang bertujuan untuk menyeimbangkan distribusi data dimana pembagian data pada jenis kepribadian memiliki penyebaran yang tidak seimbang seperti pada Tabel 1 dimana *traits Openness* yang memiliki perbandingan kelas binari 2.4(yes) : 1(no) dan Tabel 2 dimana *traits Extraversion* memiliki perbandingan kelas binari 1(yes) : 2.9(no). Teknik *Resampling* yang digunakan adalah *Under-sampling* dan *Over-sampling*.

Implementasi *deep learning* menggunakan 4 arsitektur yaitu MLP, LSTM, GRU, dan CNN 1D. Kemudian peneliti mencoba menggabungkan arsitektur LSTM dan CNN 1D. Peneliti melakukan serangkaian skenario untuk mendapatkan akurasi prediksi tertinggi untuk masing-masing arsitektur. Uji coba dilakukan dengan melakukan penambahan proses *Resampling*. Library Python yang digunakan adalah Keras dan Theano sebagai *backend*.

Tabel 3 di bawah merupakan rincian skenario percobaan yang akan dilakukan pada *machine learning* dan *deep learning*.

Machine Learning								
Skenario	Feature			Feature Selection		Resampling		
	LIWC	SPLICE	SNA	Tidak	Ya	Tanpa Resampling	Under-sampling	Over-sampling
1	✓			✓		✓		
2	✓			✓			✓	
3	✓			✓				✓
4	✓				✓	✓		
5	✓				✓		✓	
6	✓				✓			✓
7		✓		✓		✓		
8		✓		✓			✓	
9		✓		✓				✓
10		✓			✓	✓		
11		✓			✓		✓	
12		✓			✓			✓
13			✓	✓		✓		
14			✓	✓			✓	
15			✓	✓				✓
16			✓		✓	✓		
17			✓		✓		✓	
18			✓		✓			✓
Deep Learning								
Skenario	Resampling							
	Tanpa Resampling						Under-sampling	Over-sampling
19	✓							
20							✓	
21								✓

Tabel 3 Skenario percobaan *machine learning* dan *deep learning*.

4 Hasil Klasifikasi

Hasil seluruh klasifikasi *machine learning* dan *deep learning* ditunjukkan pada Tabel 4, 5, 6, dan 7. Peneliti hanya menampilkan algoritma dan arsitektur dengan akurasi tertinggi pada setiap *traits* dengan mencantumkan skenario yang digunakan di bawah setiap hasil akurasi.

Tabel 4 dengan menggunakan dataset myPersonality dan implementasi *machine learning* menunjukkan hasil akurasi tertinggi didominasi oleh skenario 1 dan 4 yang muncul sebanyak 4 kali. Skenario 1 dan 4 menggunakan fitur LIWC dan tanpa melalui proses *resampling*. Skenario 1 tanpa melalui proses *feature selection* dan skenario 4 melalui proses *feature selection*. Akurasi tertinggi didapatkan dari algoritma SVM dan Logistic Regression dengan 70.40% dan rata-rata akurasi tertinggi didapatkan dari

algoritma LDA dengan 63.04%. *Traits Openness* (OPN) memiliki rata-rata akurasi tertinggi diantara *traits* lainnya dengan 68.80%.

Tabel 5 dengan menggunakan dataset *Manual Gathering* dan implementasi *machine learning* menunjukkan hasil akurasi tertinggi kembali didominasi oleh skenario 1 dan 4. Akurasi tertinggi didapatkan dari algoritma LDA dengan 79.33% dan rata-rata akurasi tertinggi didapatkan dari algoritma SVM dengan 67.20%. *Traits Extraversion* (EXT) memiliki rata-rata akurasi tertinggi diantara *traits* lainnya dengan 75.87%.

Tabel 6 dengan menggunakan dataset myPersonality dan implementasi *deep learning* menunjukkan hasil akurasi tertinggi didominasi oleh skenario 21. Skenario 21 menggunakan proses *Resampling* dengan teknik *Under-sampling*. Akurasi tertinggi didapatkan oleh

arsitektur MLP dengan 79.49% dan rata-rata akurasi tertinggi didapatkan dari arsitektur MLP dengan 70.78%. *Traits Openness* (OPN) memiliki rata-rata akurasi tertinggi diantara *traits* lainnya dengan 74.10%.

Tabel 7 dengan menggunakan dataset *Manual Gathering* dan implementasi *deep learning* menunjukkan hasil akurasi tertinggi didominasi oleh skenario 21. Akurasi tertinggi didapatkan oleh arsitektur MLP dan LSTM+CNN 1D dengan 93.33% dan rata-rata akurasi tertinggi didapatkan oleh arsitektur LSTM+CNN 1D dengan 74.17%. *Traits Extraversion* (EXT) memiliki rata-rata akurasi tertinggi diantara *traits* lainnya dengan 83.33.

Untuk rata-rata akurasi setiap algoritma *machine learning* menunjukkan keseimbangan hasil akurasi pada kedua dataset. Namun pada implementasi *deep learning*, hasil rata-rata setiap arsitektur cukup berbeda. Sedangkan untuk *traits*, nilai rata-rata *traits Extraversion* jauh mengungguli nilai rata-rata *traits* lainnya. Hasil dari percobaan yang menggunakan implementasi *deep learning* secara rata-rata dapat mengungguli hasil yang diperoleh hanya dengan menggunakan implementasi *machine learning*. Meskipun begitu, tidak terdapat algoritma *classifier* atau arsitektur *deep learning* yang dapat menghasilkan akurasi tertinggi untuk keseluruhan *traits kepribadian Big Five Personality*.

Algoritma	Traits (Skenario)					Average
	OPN	CON	EXT	AGR	NEU	
Naive Bayes	70.00% (4)	59.20% (14)	68.80% (1)	56.40% (8)	54.40% (1)	61.76%
SVM	70.40% (4)	56.00% (4)	61.60% (4)	56.80% (12)	60.40% (4)	61.04%
Logistic Regression	70.40% (1)	54.40% (3)	68.40% (1)	53.60% (5)	60.40% (4)	61.44%
Gradient Boosting	63.20% (1)	56.40% (5)	68.00% (13)	63.20% (6)	59.20% (16)	62%
LDA	70.00% (16)	58.40% (14)	68.00% (16)	58.00% (7)	60.80% (1)	63.04%
Average	68.80%	56.88%	66.96%	57.60%	59.04%	

Table 4 Hasil klasifikasi *machine learning* dengan menggunakan dataset myPersonality. Angka di dalam tanda kurung pada setiap *traits* menunjukkan nomor skenario pada Tabel 3.

Algoritma	Traits (Skenario)					Average
	OPN	CON	EXT	AGR	NEU	
Naive Bayes	60.67% (1)	62.67% (1)	73.33% (1)	53.33% (2)	70.00% (4)	64.00%
SVM	64.67% (4)	65.33% (1)	76.00% (1)	60.67% (12)	69.33% (1)	67.20%
Logistic Regression	65.33% (7)	66.67% (11)	74.67% (4)	59.33% (5)	66.67% (1)	66.53%
Gradient Boosting	67.33% (1)	62.67% (1)	76.00% (4)	58.67% (7)	66.67% (1)	66.26%
LDA	60.00% (4)	67.33% (1)	79.33% (1)	60.67% (3)	66.67% (4)	66.80%
Average	63.60%	64.93%	75.87%	58.53%	67.87%	

Tabel 5 Hasil klasifikasi *machine learning* dengan menggunakan dataset *Manual Gathering*. Angka di dalam tanda kurung pada setiap *traits* menunjukkan nomor skenario pada Tabel 3.

Arsitektur	Traits (Skenario)					Average
	OPN	CON	EXT	AGR	NEU	
MLP	79.31% (21)	59.62% (20)	78.95% (21)	56.52% (21)	79.49% (21)	70.78%
LSTM	68.00% (19)	52.00% (19)	58.00% (19)	56.52% (21)	58.62% (20)	58.63%
GRU	68.00% (19)	62.00% (19)	58.00% (19)	65.22% (21)	64.00% (19)	63.44%
CNN 1D	79.31% (21)	50.00% (20)	60.94% (20)	67.39% (21)	61.54% (21)	63.84%
LSTM+CNN 1D	75.86% (21)	57.69% (20)	71.05% (21)	50.00% (20)	58.97% (21)	62.71%
Average	74.10%	56.26%	65.39%	59.13%	64.52%	

Tabel 6 Hasil klasifikasi *deep learning* dengan menggunakan dataset myPersonality. Angka di dalam tanda kurung pada setiap *traits* menunjukkan nomor skenario pada Tabel 3.

Arsitektur	Traits (Skenario)					Average
	OPN	CON	EXT	AGR	NEU	
MLP	66.67% (21)	64.00% (21)	93.33% (21)	70.37% (21)	75.00% (21)	73.87%
LSTM	67.50% (20)	64.00% (21)	70.00% (19)	66.67% (21)	75.00% (21)	68.63%
GRU	63.33% (19)	61.76% (20)	73.33% (21)	59.38% (20)	76.67% (19)	66.89%
CNN 1D	76.19% (21)	68.00% (21)	86.67% (21)	63.33% (19)	75.00% (21)	73.84%
LSTM+CNN 1D	67.50% (20)	66.67% (19)	93.33% (21)	63.33% (19)	80.00% (21)	74.17%
Average	68.24%	64.89%	83.33%	64.62%	76.33%	

Tabel 7 Hasil klasifikasi *deep learning* dengan menggunakan dataset *Manual Gathering*. Angka di dalam tanda kurung pada setiap *traits* menunjukkan nomor skenario pada Tabel 3.

Dalam penelitian ini, peneliti telah melakukan percobaan terhadap prediksi kepribadian berdasarkan model kepribadian *Big Five Personality Traits*. Klasifikasi dilakukan dengan implementasi *machine learning* dan *deep learning* dengan melakukan beberapa skenario percobaan.

Implementasi *machine learning* menggunakan 5 algoritma yaitu Naive Bayes, SVM, *Logistic Regression*, *Gradient Boosting*, dan LDA serta menggunakan 3 jenis fitur yaitu LIWC, SPLICE, dan SNA. Evaluasi model yang digunakan adalah *10-fold cross validation*. Skenario percobaan pada *machine learning* terdiri dari penggunaan 2 dataset, *feature selection* dan *resampling*. Skenario percobaan dengan menggunakan dataset myPersonality didapatkan akurasi tertinggi 70.40%

menggunakan algoritma SVM dan *Logistic Regression* pada *traits Openness* (OPN) dengan menggunakan fitur LIWC. SVM melalui proses *feature selection* dan *Logistic Regression* tanpa melalui proses *feature selection* serta kedua algoritma tersebut tanpa dilakukan proses *resampling*. Skenario percobaan dengan menggunakan dataset *Manual Gathering* didapatkan akurasi tertinggi 79.33% dengan menggunakan algoritma LDA pada *traits Extraversion* (EXT) dengan menggunakan fitur LIWC, tanpa dilakukan proses *feature selection* dan tanpa dilakukan proses *resampling*.

Hasil percobaan pada *machine learning* membuktikan bahwa penggunaan algoritma LDA memiliki rata-rata akurasi pada kedua dataset yang paling tinggi tetapi tidak berbeda jauh dengan algoritma lainnya dengan selisih

maksimal 2.04% dan selisih minimal 0.79%. Kemudian 85 fitur LIWC tanpa dilakukan proses *feature selection* memiliki akurasi yang lebih tinggi daripada fitur 74 fitur SPLICE dan 7 fitur SNA walaupun dilakukan percobaan pada 2 dataset yang berbeda. Peneliti juga mencoba melakukan kombinasi fitur LIWC, SPLICE, dan SNA tetapi tidak berhasil meningkatkan akurasi. Teknik *resampling* pada implementasi *machine learning* juga tidak dapat meningkatkan akurasi.

Implementasi *deep learning* menggunakan 4 arsitektur yaitu MLP, LSTM, GRU, dan CNN 1D serta peneliti mencoba menggabungkan arsitektur LSTM dengan CNN 1D. Skenario percobaan pada *deep learning* terdiri dari 2 dataset dan *resampling*. Skenario percobaan dengan menggunakan dataset myPersonality didapatkan akurasi tertinggi 79.49% dengan menggunakan arsitektur MLP pada *traits Openness* (OPN) dan dilakukan proses *resampling* dengan teknik *Under-sampling*. Skenario percobaan dengan menggunakan dataset *Manual Gathering* didapatkan akurasi tertinggi 93.33% dengan menggunakan arsitektur MLP dan LSTM+CNN 1D pada *traits Extraversion* (EXT) dan dilakukan proses *resampling* dengan teknik *under-sampling*.

Hasil percobaan pada *deep learning* membuktikan bahwa penggunaan arsitektur MLP memiliki rata-rata akurasi pada kedua dataset yang paling tinggi dengan selisih minimal 3.48% dan selisih maksimal 8.69%. Teknik *resampling* juga terbukti mampu meningkatkan akurasi secara signifikan hampir pada seluruh skenario percobaan.

5 Kesimpulan

Hasil dari penelitian ini menunjukkan penggunaan *deep learning* dapat meningkatkan hasil akurasi dengan menerapkan arsitektur dan proses yang tepat. Walaupun begitu, hasil yang didapatkan masih tergolong rendah untuk beberapa *traits*. Alasan yang paling kuat menurut peneliti adalah jumlah dataset yang digunakan dalam penelitian ini masih terlalu kecil. Namun, hasil yang didapatkan oleh kedua implementasi dalam penelitian ini dapat mengungguli hasil dari penelitian sebelumnya yang menggunakan dataset yang sama.

Untuk itu, pada penelitian selanjutnya, peneliti berencana untuk mendapatkan dataset lebih banyak lagi dari myPersonality. Peneliti juga berupaya untuk menggunakan algoritma XGBoost dan kembali menerapkan implementasi

deep learning dengan kombinasi arsitektur dan proses yang tepat untuk meningkatkan dan mengembangkan sistem prediksi ini.

Referensi

- A. James O'Malley, Peter V. Marsden (2008) *The analysis of social networks*. Health Services and Outcomes Research Methodology, (8), 222–269.
- Albert Wijaya, Nathanael Febrianto, Irwan Prasetya, and Derwin Suhartono (2016) *Sistem Prediksi Kepribadian "The Big Five Traits" Dari Data Twitter*. Bina Nusantara University, Jakarta, Indonesia.
- Firoj Alam, Evgeny A. Stepanov, and Giuseppe Riccardi (2013) *Personality Traits Recognition on Social Network – Facebook*. WCPR (ICWSM-13), Cambridge, MA, USA.
- Golnoosh Farnadi, Susana Zoghbi, Marie-Francine Moens, and Martine De Cock (2013) *How Well Do Your Facebook Status Updates Express Your Personality?*. Conference on Machine Learning, Nijmegen, The Netherlands.
- H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, . . . Lyle H. Ungar (2013) *Personality, Gender, and Age in the Language of Social Media: The Open Vocabulary Approach*. PLOS ONE, 8, e73791.
- James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn (2015). *The development and Psychometric Properties of LIWC2015*. University of Texas, Austin, Texas.
- Kevin C. Moffit, Justin S. Giboney, E. Ehrhardt, Judee K. Burgoon, and Jay F. Nunamaker (2012) *Structured Programming for Linguistic Cue Extraction (SPLICE)*. Report of the HICSS-45
- Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium, 103-108.
- Michal Kosinski, Sandra C. Matz, Samuel D. Gosling, Vesselin Popov, and David Stillwell (2015) *Facebook as a Social Science Research Tool: Opportunities, Challenges, Ethical Considerations and Practical Guidelines*. American Psychologist.
- Navonil Majumder, Soujanya Poria, Alexander Gelbukh, Erik Cambria (2017) *Deep learning-Based Document Modeling for Personality Detection from Text*. IEEE Intelligent Systems, 32(2), 74-79.

- Veronica Ong, Anneke D. S. Rahmanto, Williem, and Derwin Suhartono (2017). *Personality Prediction Based on Twitter Information in Bahasa*. Internetworking Indonesia Journal, 9(1), 65-70.
- Yoram Bachrach, Michal Kosinski, Thore Graepel, Pushmeet Kohli, and David Stillwell (2012) *Personality and Patterns of Facebook Usage*. ACM Web Science Conference. *Proceedings of the ACM Web Science Conference*, 36–44.