Sistem Prediksi Kepribadian Big Five Personality

Berdasarkan Data Pengguna Facebook

|  |  |  |
| --- | --- | --- |
| Tommy Tandera  Computer Science Department  School of Computer Science  Bina Nusantara University  Indonesia  tommy.tandera@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 Prasetio  Computer Science Department  School of Computer Science  Bina Nusantara University  Indonesia  yenlina@binus.edu |

Abstract

The usage of social networks has now reached its peak. Various information shared widely through social media such as Facebook. Information about users and their statuses is such an important asset for research in the field of behavioral learning and human personality. Similar researches have been conducted in this field and continue to grow to date. This study attempts to build a system that can predict a person’s personality based on Facebook user information. Personality model that used in this research is Big Five Model Personality. While other previous research using older machine learning algorithm in building their model, this research try to implement some deep learning architectures to see the comparison by doing comprehensive analysis method through the accuracy result. The results shown in this study succeeded to outperform the accuracy of previous similar research with the current highest accuracy of 93.33% acquire using deep learning architecture.

1 Pendahuluan

Social media has become the most used communication and interaction tool between people over the past few years. In the era where almost all human beings have their own smartphones, direct interaction between people almost rarely happens. So, it is quite difficult to recognize and get to know the personality of a person. However, this is totally different from what happens in social media. Facebook has the largest users reaching 1.8 billion users with aroung 800 million users spending about 40 minutes a day using Facebook (Bachrach et al., 2012). Facebook users generally express their feelings and opinions in their user feed. Although Facebook is currently more widely used to share photos and videos, this research will be focus on users’ linguistic aspect which is their statuses. Various studies in the field of psychology show that there is a correlation between personality and the linguistic behavior of a person. This correlation can be effectively analyzed and illustrated using NLP approach. Therefore this research goal is to build a prediction system that can automatically predict an user personality based on their activity in Facebook.

This prediction system will be built using the Big Five Personality model. There are several other personality models used in related study such as MBTI (Myers-Briggs Type Indicator) or DISC. However, after some considerations and literature review process, Big Five Personality is chosen by the reason it’s the most popular and precise in telling someone’s personality traits. Traits in this model consist of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

The corpus in this study will be divided by two datasets. First dataset consists of 250 users with around 10,000 statuses obtained from myPersonality project sample data and the other dataset of 150 users collected manually. Prediction system will be built using some linguistic features with different approach. The first is using closed vocabulary that includes some features such as LIWC and SPLICE. SNA (Social Network Analysis) also included in the process because all the features score provided by myPersonality dataset. All features in the first approach specifically used in the older machine learning algorithm implementation. The second approach is using open vocabulary approach which will be word embedding features specifically used in deep learning technique implementation. Similar research using machine learning older algorithm which is we’ll also use in this research has been widely used before, but the implementation of deep learning in this field of research still hardly to find. Therefore, this research will also conduct the implementation of deep learning to see whether it can boost the result of the prediction system. The best classifier and features will be based on the accuracy result and will be used as the model for the final personality prediction system.

**2 Related Works**

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 penelitan 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**

* 1. Dataset

The dataset used in this study is divided into two parts. The first dataset obtained from myPersonality (Kosinski et al., 2015) as many as 250 datasets of Facebook users with approximately 10,000 statuses that have been given labeling personality based on the Big Five Personality Traits model. The distribution of the myPersonality dataset based on the personality type is presented in Table 1 below.

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

Table 1 Distribution of myPersonality dataset

The second dataset is the status of 150 Facebook user datasets collected manually. Facebook API Graph is utilized in the process of collecting the dataset. Personality labeling is then done by manually entering the user posts into applymagicsauce app. Table 2 is the result of dataset distribution after being labeled based on Big Five Personality Traits model.

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

Tabel 2 Distribution of Manual Gathering dataset

* 1. Fitur-fitur yang digunakan

This study will use several features to see the comparison of results and capabilities between them. The main reason is to investigate the suitability and performance of this various features for personality modeling. The features used are differentiated for each of the learning implementations. For machine learning implementation, we use linguistic feature with closed-vocabulary approach. Closed vocabulary is a feature based on the number of words content in accordance with predefined features. For this approach, we used linguistic features such as LIWC (Pennebaker, 2015) and SPLICE (Moffit et al., 2012). LIWC used in this study is LIWC2015 version which has 85 features that have been developed from LIWC2007 version. In this study all LIWC features will be used.

SPLICE is a linguistic feature created by Moffit et al and has been used in several studies in this field. In this study there are 74 features of SPLICE that will be used.

In addition to the above linguistic features, this research will also utilize the use of Social Network Analysis features provided by the myPersonality dataset in the form of detailed information about a user's friendship network. For complete information on this feature can be seen in (O'Malley & Marsden, 2008).

In contrast to the implementation of machine learning, implementation of deep learning was done separately by using linguistic features of open vocabulary approach. Open vocabulary does not require predefined features. This approach will perform an automatic exploration of the dataset used to find the relationship between the uses of words with personality. The actual technique that used in this study is word embedding using GloVe. Previous studies that have made comparisons between these two linguistic feature approaches have been done in (Schwartz et al., 2013).

* 1. Preprocessing

All data that has been collected in this research will go through the preprocessing stage before build the classification model. Pre-preprocessing steps are removing URLs, remove symbols, remove names, remove spaces, lower case text, stemming, and remove stopwords.

Especially for status with Indonesian language, additional preprocessing process is done manually by replacing slang words or non-standard words from the status first to then proceed to the translation into English.

* 1. Klasifikasi Model

Implementation of machine learning using 5 different algorithms, namely Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Gradient Boosting, and Linear Discriminant Analysis (LDA). For model validation, researchers used a 10-fold cross validation technique using Python libraries. 10-fold cross validation divides 10% dataset into data testing and 90% dataset as training data in turn.

Researchers conducted a series of tests with various scenarios to see how the algorithm accuracy results in predicting the personality type. Testing is done by adding some additional processes to improve accuracy. The first process is Features Selection that tries to filter or remove the features used that are considered to have a low correlation to the traits of the personality. The next process is to do a resampling process that aims to balance the distribution of data where the data distribution on the personality type has an unbalanced distribution as in Table 1 where Openness traits have a comparison of binary classes 2.4 (yes): 1 (no) and Table 2 where Traits Extraversion has a binary class comparison of 1 (yes): 2.9 (no). The resampling technique used is Under-sampling and Over-sampling.

Implementation of deep learning using four architectures, namely MLP, LSTM, GRU, and CNN 1D. Then the researchers tried to combine LSTM and CNN 1D architecture as an additional architecture. Researchers conducted a series of scenarios to obtain the highest prediction accuracy for each architecture. The test is done by adding the resampling process. The Python library used is Keras and Theano as the backend.

Table 3 below is a breakdown of experimental scenarios to be performed on machine learning and 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*.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Algorithm** | ***Traits* (Scenario)** | | | | | **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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Algorithm** | ***Traits* (Scenario)** | | | | | **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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Architecture** | ***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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Arschitecture** | ***Traits* (Scenario)** | | | | | **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 Prasetia, 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.