

Penerapan Linguistic Inquiry and Word Count dan Random Forest Dalam Klasifikasi Personality Berdasarkan Data Posting Twitter Sehingga Dapat Ditentukan Gaya Belajar yang Sesuai

Cristine Ferlly Wiyanto, Henry Novianus Palit, Alvin Nathaniel Tjondrowiguno
Program Studi Informatika Fakultas Teknologi Industri Universitas Kristen Petra

Jl. Siwalankerto 121 – 131 Surabaya 60236
Telp. (031) – 2983455, Fax. (031) - 8417658

E-mail: ferlly.wiyanto@gmail.com, hnpalit@petra.ac.id, alvin.nathaniel@petra.ac.id

ABSTRAK

Big Five Personality merupakan model kepribadian yang kuat untuk memahami hubungan antara kepribadian dan berbagai perilaku akademik. Kepribadian siswa sangat penting untuk belajar dan berpotensi menentukan prestasi akademik maupun gaya belajar mereka. Namun, tidak semua pelajar memiliki pengetahuan, kepribadian, dan gaya belajar yang sama dimana kriteria-kriteria ini mempengaruhi dalam pembelajaran. Tetapi untuk mengetahui hal tersebut, biasanya kita menggunakan tes online sedangkan tes online membutuhkan waktu yang lama. Di dalam penelitian ini, sebuah sistem dibuat untuk menentukan personality dan *learning style* secara otomatis berdasarkan data posting Twitter.

Metode yang digunakan dalam penelitian ini adalah LIWC atau Linguistic Inquiry and Word Count dan Random Forest. Random Forest dipilih karena metode ini dapat melakukan klasifikasi terhadap ketidakseimbangan kelas dimana dalam melakukan klasifikasi *Big Five personality* dari data teks, tidak semua data tersebut memiliki jumlah yang sama antara kelima *personality* (*extraversion, agreeableness, openness, conscientiousness, and neuroticism*).

Hasil akurasi Random Forest untuk dataset *balanced* dan *imbalanced* tidak terlalu signifikan, seperti *personality CON imbalanced* memiliki akurasi sebesar 0.499 sedangkan *CON balanced* memiliki akurasi sebesar 0.502 atau juga seperti *personality NEU imbalanced* memiliki akurasi sebesar 0.502 sedangkan *NEU balanced* memiliki akurasi sebesar 0.519. Sedangkan hasil *learning style* dapat ditentukan dari *Big Five Personalities* dengan nilai rata-rata korelasi Kendall Tau sebesar 0.21. Menurut survei kecocokan terhadap responden, responden merasa web eksternal lebih cocok dengan nilai rata-rata kecocokan responden dengan hasil dari web eksternal 4.5 untuk *Big Five Personality* dan 4 untuk hasil *learning style*. Sementara itu, untuk hasil dari program, rata-rata yang didapat untuk *Big Five Personality* adalah 3 sedangkan untuk *learning style* memiliki nilai rata-rata 3.25.

Kata Kunci: *Big Five Personality, Random Forest, LIWC, kepribadian, akademik, tes online*

ABSTRACT

Big Five Personality is a powerful personality model for understanding the relationship between personality and various academic behaviors. Students' personality is very important for learning and has the potential to determine their academic achievement and learning style. However, not all students have the same knowledge, personality, and learning styles where these criteria affect learning. To find out, we usually use online tests and

it takes a long time. In this study, a system was created to determine personality and learning style automatically based on Twitter post data.

The method used in this research is LIWC or Linguistic Inquiry and Word Count and Random Forest. Random Forest was chosen because this method can classify class imbalances where in classifying the Big Five personalities from text data, not all of the data have the same number of personalities (extraversion, agreeableness, openness, conscientiousness, and neuroticism). The data text that will be used is data text from social media, namely Twitter with a total data of 9546 data.

The results of Random Forest accuracy for balanced and imbalanced datasets are not very significant, such as the imbalanced CON personality has an accuracy of 0.499 while the balanced CON has an accuracy of 0.502 or also the imbalanced NEU personality has an accuracy of 0.502 while the balanced NEU has an accuracy of 0.519. While the results of learning style can be determined from the Big Five Personalities with an average Kendall Tau correlation value of 0.21. According to the compatibility survey of the respondents, respondents felt that the external web was more suitable with the average value of the respondent's suitability with the results of the external web of 4.5 for Big Five Personality and 4 for learning style results. Meanwhile, for the results of the program, the average obtained for the Big Five Personality is 3 while for the learning style it has an average value of 3.25.

Keywords: *Big Five Personality, Random Forest, LIWC, personality, academic, online test*

1. PENDAHULUAN

Pelajar dapat menerima materi dengan baik apabila mereka tahu bagaimana cara atau gaya belajarnya. Jika gaya belajar tidak cocok maka pelajar bisa kehilangan motivasi. Gaya belajar mengacu pada cara yang disukai individu untuk memproses informasi baru untuk pembelajaran yang efisien [9]. Gaya belajar adalah tentang bagaimana siswa belajar dari apa yang mereka telah pelajari [8]. Proses belajar berbeda untuk setiap individu. Walaupun di dalam lingkungan pendidikan yang sama, pembelajaran antar siswa tidak sama. Penelitian telah menunjukkan bahwa setiap individu menunjukkan pendekatan yang berbeda dalam proses pembelajaran dan strategi [11]. Hal ini mungkin terkait dengan perbedaan latar belakang siswa, minat, ambisi, tingkat motivasi, dan gaya belajar [5].

Kepribadian siswa sangat penting untuk belajar dan berpotensi menentukan prestasi akademik maupun gaya belajar mereka [16]. Gaya belajar mengacu pada preferensi individu dalam memperoleh

informasi, keterampilan, sikap atau pengetahuan melalui pengalaman atau belajar. Sejumlah penelitian menunjukkan bahwa kepribadian merupakan prediktor yang baik dari keberhasilan akademik, terutama pada konteks pendidikan tinggi [4]. Oleh karena itu, setiap individu memiliki gaya belajar mereka sendiri dan ditambah dengan kecerdasannya akan mengarahkan setiap individu pada prestasi akademik [13]. Menurut [3], tipe kepribadian dan gaya belajar memiliki hubungan yang erat. Di dalam [12] dan [16], *Big Five Personality* merupakan model kepribadian yang kuat untuk memahami hubungan antara kepribadian dan berbagai perilaku akademik.

Menurut [2], dalam dekade terakhir, psikolog kepribadian, bersama dengan ilmuwan komputer, berusaha untuk menggali kepribadian *Big Five Personality* berdasarkan kegiatan media sosial seperti Twitter. Ide dasar dalam penelitian ini adalah kepribadian tercermin dalam cara orang menulis dan bahwa apa yang ditulis mencakup informasi tentang karakteristik dari kepribadian penulis. Selain itu, melakukan kuesioner dapat menimbulkan bias pada hasil dan juga memakan waktu yang lama. Oleh karena itu, media sosial merupakan sumber yang paling tepat untuk dijadikan dataset karena media sosial menampung seluruh perasaan, cerita dari pengguna. Sehingga di dalam penelitian ini, sosial media yang dipakai adalah Twitter. Menurut [6], Twitter lebih banyak digunakan untuk berbagi informasi personal sementara media sosial lainnya seperti Facebook lebih banyak digunakan untuk menghubungkan dengan kelompok tertentu dan berbagi konten yang lebih mendalam.

Di dalam penelitian [17], mereka telah melakukan penelitian untuk memprediksi kepribadian Big Five dari data posting di Facebook. Mereka menggunakan LIWC (Linguistic Inquiry and Word Count) and SPLICE (Structured Programming for Linguistic Cue Extraction) untuk linguistic feature. Di dalam studi ini, peneliti membandingkan antara metode yang tradisional seperti Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Gradient Boosting, and Linear Discriminant Analysis (LDA) dengan deep learning algorithm seperti MLP (Multi-Layer Perceptron), LSTM (Long Short Term Memory), GRU (Gated Recurrent Unit), dan CNN 1D (1-Dimensional Convolutional Neural Network). Dalam penelitian [1] Random Forest menggabungkan ratusan atau ribuan *decision trees* yang digunakan untuk melatih masing-masing set yang berbeda, memberikan batasan pada *features* yang dapat dipakai maupun menggabungkannya untuk mendapatkan yang prediksi yang lebih akurat dan stabil. Hal ini terbukti dari hasil yang dicapai dari penelitian ini. Hasil Random Forest saat memprediksi kepribadian MBTI dapat mencapai 100% jika dibandingkan dengan metode learning yang lain seperti Logistic Regression, KNN neighbor dan Support Vector Machine (SVM). Pada penelitian yang dilakukan [7], mereka menggunakan *linguistic features* yang dipakai terbagi menjadi dua kategori, yaitu *style-based features* seperti *part-of-speech* dan *type token ratio* dan *content-based features* seperti LIWC, N-grams, Latent Semantic Analysis. Untuk proses *machine learning*, metode yang akan dipakai ada tiga yaitu Logistic Regression, Random Forest dan Support Vector Machine. Diantara ketiga metode tersebut, dalam pengenalan kepribadian, metode Random Forest bekerja dengan baik dalam banyak kasus [7]. Oleh karena itu, pada skripsi ini akan menggunakan LIWC dan Random Forest dengan fokus pengerjaannya lebih kepada klasifikasi dari hasil *personality*. Dengan hasil klasifikasi tersebut, digunakan untuk menentukan gaya belajar yang sesuai dengan cara mengalikan hasil numerik *Big Five personality* dengan tabel

korelasi antara *personality* dengan gaya belajar berdasarkan penelitian [16].

2. PENELITIAN SEBELUMNYA

Masalah yang diangkat dalam penelitian [17] adalah membangun sistem prediksi yang secara otomatis dapat memprediksi kepribadian pengguna berdasarkan kepribadiannya menurut *Big Five Personality* yang dilihat dari aktivitas yang dilakukan di Facebook. Hasil *Big Five Personality* yang dihasilkan merupakan multi label. Metode yang digunakan ialah LIWC atau Linguistic Inquiry and Word Count dan SPLICE atau Structured Programming for Linguistic Cue Extraction untuk *linguistic extraction* dan untuk *machine learning*nya menggunakan metode yang tradisional seperti Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Gradient Boosting, dan Linear Discriminant Analysis (LDA) dengan deep learning algorithm seperti MLP (Multi-Layer Perceptron), LSTM (Long Short Term Memory), GRU (Gated Recurrent Unit), dan CNN 1D (1-Dimensional Convolutional Neural Network). Skenario eksperimental dilakukan dengan dua cara, yang pertama ialah menggunakan dataset dan yang kedua ialah dengan *resampling*. Hasil yang didapat ialah pada machine learning tradisional membuktikan bahwa algoritma LDA memiliki akurasi rata-rata tertinggi yaitu dalam dataset myPersonality dan algoritma SVM memiliki akurasi rata-rata tertinggi untuk dataset yang dikumpulkan secara manual namun tidak jauh berbeda dengan algoritma lainnya. Sedangkan dalam *deep learning*, akurasi tertinggi sebesar 79,49% untuk *Neuro* diperoleh dengan menggunakan dataset dengan metode MLP sedangkan untuk data yang dikumpulkan secara manual menunjukkan bahwa akurasi tertinggi adalah 93,33% untuk *Extra*.

Masalah yang diangkat dalam penelitian [7] adalah peneliti ingin melakukan pendekatan pengenalan kepribadian *Big Five personality* yang dapat ditafsirkan berdasarkan teks yang ada di media sosial (Weibo). Hasil *Big Five Personality* yang dihasilkan merupakan multi label. Metode yang digunakan ialah *text features* seperti LIWC, N-Gram, Latent Semantic Analysis (LSA) dengan metode 3 metode *machine learning* yaitu Logistic Regression, Random Forest dan Support Vector Machine. Hasil yang didapat ialah, dari ketiga metode *machine learning* yang dipakai, metode klasifikasi Random Forest lebih unggul dibanding dengan dua metode lainnya dengan tingkat akurasi diantara 0.692 untuk kepribadian *Neuro*, 0.708 untuk *Extra*, 0.72 untuk *Cons*.

Masalah yang diangkat dalam penelitian [1] adalah sulitnya memilih karyawan yang tepat karena banyaknya calon. Secara tradisional, sebuah perusahaan akan mengatur sesi wawancara dengan calon kandidat untuk mengetahui kepribadian. Namun, prosedur ini terkadang menuntut ekstra waktu karena jumlah wawancara lebih sedikit dari jumlah pencari kerja. Metode yang digunakan adalah Random Forest, Logistic Regression, KNN Neighbor and Support Vector Machine (SVM) serta word2vec untuk *linguistic processing*-nya. Dengan pembagian dataset *training testing* masing-masing 90% dan 10%. Hasil yang dicapai pada penelitian sebelumnya adalah dapat memprediksi kepribadian dengan menggunakan data sosial media, dan model terbaik adalah Random Forest dengan rata-rata 100% dalam keempat dimensi kepribadian MBTI. Perbedaan penelitian yang dilakukan dengan penelitian sebelumnya ada pada jenis kepribadian yang digunakan. Pada penelitian sebelumnya, tipe kepribadian yang digunakan adalah MBTI atau *Myers-Briggs Type Indicator* sedangkan pada skripsi ini menggunakan *Big Five Personality*. Selain itu, di dalam skripsi ini menggunakan LIWC sebagai *linguistic processing*-nya.

3. DATASET

3.1 Dataset Essay

Dataset *essay* ini milik James Pennebaker and Laura King. James Pennebaker merupakan ahli psikologi asal Amerika. Data essay ini merupakan hasil penelitian yang dilakukan oleh keduanya di tahun 1999. Data ini memuat sekitar 2400 *essay* yang sudah diberikan label *personality*. *Personality* yang digunakan ialah *Big Five Personality*. Data *essay* ini memiliki 2467 buah data esai yang anonim yang sudah diberi tag EXT untuk *extraversion*, NEU untuk *neuroticism*, AGR untuk *agreeableness*, CON untuk *conscientiousness*, and OPN untuk *openness*.

3.2 Data Posting Twitter

Data posting Twitter yang digunakan berasal dari situs Kaggle dan akan diambil secara manual dengan *library* tweepy. Untuk data yang didapat dari Kaggle terdapat 9546 dari total 9916 record *tweet* meskipun terdapat beberapa ID yang sama [18]. Akan tetapi data tersebut akan dikelompokkan berdasarkan pengguna sehingga hanya terdapat 247 *record* data.

4. METODE

Metode yang digunakan terdiri dari dua yaitu LIWC dan *Random Forest*. LIWC adalah alat analisis teks yang digunakan untuk mengevaluasi sifat psikologis dari bahasa [15]. Kemudian hasil tersebut dimasukkan ke dalam *Random Forest*. Alur pengerjaan dapat dilihat di Gambar 1 dan Gambar 2. Kemudian juga terdapat metode SVM yang digunakan untuk melakukan tes.

4.1 LIWC

LIWC (*Linguistic Inquiry dan Word Count*) adalah alat analisis teks yang digunakan untuk mengevaluasi sifat psikologis dari bahasa [15]. Menurut website LIWC [10], program LIWC akan melakukan analisis teks dimana di dalamnya juga memiliki sekelompok kamus bawaan. LIWC membaca teks verbal tertulis atau transkripsi yang telah disimpan dalam bentuk digital yang dapat dibaca komputer (seperti file teks). Modul analisis teks mengidentifikasi dan mengkategorikan kata, sehingga nantinya sekelompok kamus akan memberi tahu modul analisis teks kata mana saja yang harus diidentifikasi dan diklasifikasi. Di dalam kamus tersebut yang dapat mendefinisikan satu atau lebih kategori kata. Setelah modul pemrosesan membaca dan menghitung semua kata dalam teks yang diberikan, modul ini menghitung persentase total kata yang cocok dengan setiap kategori kamus.

4.2 Random Forest

Random forest terdiri dari beberapa individual *decision trees* yang beroperasi sebagai ansambel. Setiap *tree* yang berada di *random forest* akan mengeluarkan prediksi kelas dimana kelas dengan suara terbanyak menjadi model untuk melakukan prediksi. Dataset yang telah disiapkan akan dimasukkan ke dalam proses *bagging* dimana dataset akan di *sampling* secara acak dan pengambilan *sampling* tersebut dapat berulang. Setelah itu *trained model* yang sudah terbentuk akan diterapkan untuk melakukan prediksi terhadap sekumpulan data baru. Prediksi yang dihasilkan akan disatukan di dalam proses *mean/mode* atau bisa disebut juga sebagai *majority voting* untuk mendapatkan nilai prediksi final.

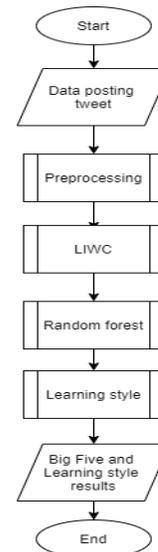
4.3 Convert Big Five Personality ke Learning Style

Setelah hasil *Random Forest* didapatkan, nilai *Random Forest* tersebut dikalikan dengan nilai korelasi dari [16]. Kemudian nilai perkalian tersebut ditambahkan berdasarkan tipe *learning style* dan hasilnya diurutkan dari yang paling tinggi hingga paling rendah.

5. PENGUJIAN

5.1 Pengujian Parameter *Big Five Personality*

Pada awal mula pengujian parameter, pengujiannya masih secara manual. Manual disini adalah mencoba satu per satu kombinasi dari parameter. Rasio training dan testing data tetap sama dengan pengujian umum yaitu 80 : 20. Terdapat empat kali rangkaian untuk pengujian parameter sebelum penggunaan *GridSearchCV*. Dalam menentukan parameter yang paling tepat dari percobaan *GridSearchCV* terdapat dua cara yang digunakan. Cara pertama dengan melihat dari nilai *rank_test_score* yang paling tinggi. Cara kedua dengan melihat jumlah selisih dari mean yang terkecil. Untuk mendapatkan jumlah selisih dari mean yang terkecil terdapat beberapa langkah. Langkah pertama melakukan selisih absolut dari *split0_test_score* terhadap *mean_test_score*. Langkah ini dilakukan dari *split0_test_score* sampai *split4_test_score*. Berikutnya, setelah mendapatkan nilai selisih absolut, hasil selisih tersebut dijumlahkan per barisnya dari *split0_test_score* sampai *split4_test_score*. Kemudian, melihat nilai minimum dari langkah sebelumnya. Memilih parameter yang nilai selisih jumlah dari proses b memiliki nilai yang sama dengan nilai minimum tetapi hasil yang didapat masih *overfitting*. Hasil dapat dilihat pada Tabel 1.

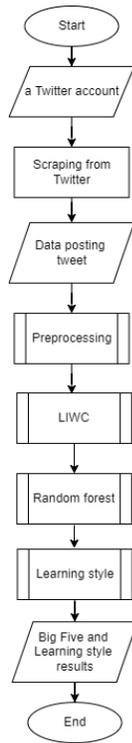


Gambar 1. Flowchart Desain Sistem Secara Keseluruhan

5.2 Pengujian Random Forest Dengan Support Vector Machine (SVM).

Support Vector Machine (SVM) juga menggunakan *GridSearchCV* untuk mendapatkan kombinasi parameter yang bagus. Untuk menentukan parameter pada *Support Vector Machine* (SVM) sama dengan cara yang digunakan dalam menentukan parameter pada pengujian *Random Forest*. Untuk dataset yang digunakan adalah dataset *imbalanced* dan *balanced*, Dataset *balanced* didapatkan dengan cara augmentasi atau menambahkan dataset *essay* milik James Pennebaker and Laura King dengan memperhatikan kesamaan OCEAN nya. Hasil tes akurasi yang didapatkan lebih tinggi daripada hasil *Random Forest*, hasil dapat dilihat pada Tabel 2. Akan tetapi, saat dilihat dari hasil *confusion*

matrix, hasil yang didapat menunjukkan bahwa model SVM gagal belajar.



Gambar 2. Flowchart Desain Sistem Secara Keseluruhan Untuk Memproses Satu Akun Twitter

5.3 Pengujian Dengan Real User

Real user akan melakukan tes online yang berasal dari pihak ketiga. Tes online tersebut adalah <https://bigfive-test.com> untuk tes *Big Five Personality* dimana tes ini dibuat berdasarkan *Standard International Personality Item Pool* atau IPIP. Sedangkan untuk tes cara belajar menggunakan <https://www.webtools.ncsu.edu/learningstyles/>, tes ini sudah diteliti untuk reabilitas dan validitasnya pada [14]. Hasil *real user* akan dibandingkan dengan web eksternal yang terlebih dahulu dibagi 1.2. Sedangkan untuk perbandingan gaya belajar atau *learning style* terdapat dua proses perbandingan. Pertama, hasil *Big Five* dari web eksternal yang dimasukkan ke dalam proses *learning style* program dibandingkan dengan hasil *learning style* program. Kedua, hasil *learning style* dari web eksternal dibandingkan dengan hasil *learning style* program.

Kedua proses perbandingan ini diimplementasikan ke dalam dua cara, dimana cara yang pertama adalah mengalikan setiap hasil *Big Five Personality* dengan nilai korelasi dari *learning style*. Untuk cara yang kedua adalah dengan cara mengalikan hasil *Big Five Personality* dengan angka yang memiliki nilai *confidence* dibawah 0.5 / 0.01 (yang memiliki satu / dua simbol * atau asterisk), dapat dilihat pada penelitian milik [16]. Hal ini dikarenakan adanya asumsi bahwa yang memiliki nilai *confidence* dibawah 0.5 dan 0.01 memiliki korelasi yang kuat sehingga hanya angka-angka yang diberi simbol * yang akan dioperasikan dengan *Big Five Personality* untuk cara kedua. Proses perbandingan *learning style*

ini menggunakan metode Kendall Tau *Correlation* dimana yang dibandingkan adalah urutan/*ranking* bukan hasil akurasi. Semakin urutan *learning style* antara *ground truth* (web eksternal) dan program sama maka nilai Kendall Tau akan semakin tinggi yang berarti hasilnya benar. Sebaliknya apabila urutan tidak sama maka hasil korelasi Kendall Tau akan kecil yang berarti hasilnya kurang cocok. Perbandingan ini merupakan cara pertama untuk memvalidasi hasil program. Selain itu, juga dilakukannya survei kecocokan dari hasil yang didapat oleh para responden dari web eksternal dan juga dari hasil program. Survei berikut adalah cara validasi kedua dari hasil yang didapat.

Para responden akan menilai dengan angka 1 sampai 5, dengan keterangan sebagai berikut: 1 - hasil sangat tidak cocok, 2 - hasil kurang cocok, 3 - hasil netral, 4 - hasil cocok, 5 - hasil sangat cocok.

Tabel 1. Hasil Ranking dan Jumlah Selisih dari GridSearchCV

		Random Forest			
		max_features	min_samples_leaf	n_estimators	test_score
EXT	mean	log2	5	50	0.501
	highest	none	10	50	0.495
NEU	mean	none	5	200	0.502
	highest	log2	1	50	0.506
AGR	mean	sqrt	5	50	0.503
	highest	sqrt	10	50	0.503
CON	mean	log2	10	100	0.499
	highest	sqrt	5	50	0.498
OPN	mean	sqrt	10	200	0.497
	highest	sqrt	10	50	0.497

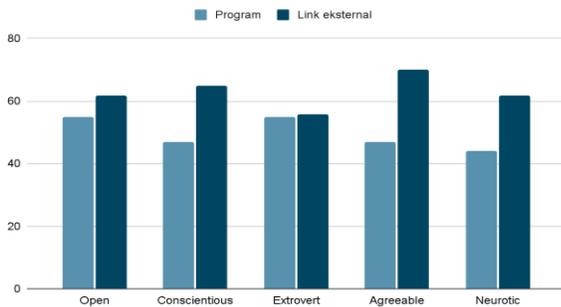
Tabel 2. Hasil dari Ranking dan Jumlah Selisih GridSearchCV

			SVM					
			C	gamma	kernel	degree	test_score	
Imbalanced	EXT	mean	10	auto	sigmoid	3	0.56	
		highest	10	auto	linear	3	0.56	
	NEU	mean	1	scale	rbf	3	0.54	
		highest	0.1	scale	sigmoid	3	0.54	
	AGR	mean	10	scale	poly	3	0.4	
		highest	10	auto	poly	3	0.48	
	CON	mean	10	scale	linear	3	0.5	
		highest	10	auto	poly	3	0.38	
	OPN	mean	1	scale	rbf	3	0.66	
		highest	10	auto	poly	3	0.66	
	Balanced	EXT	mean	10	auto	sigmoid	3	0.529
			highest	10	scale	linear	3	0.611
NEU		mean	10	auto	sigmoid	-	0.6	
		highest	1	scale	rbf	-	0.729	
AGR		mean	0.1	scale	linear	-	0.503	

	CON	highest	10	auto	linear	-	0.551
		mean	0.1	scale	linear	-	0.505
		highest	10	scale	linear	-	0.541
	OPN	mean	0.1	scale	linear	-	0.505
		highest	1	scale	rbf	-	0.752

Hasil dari responden pertama dapat dilihat pada Gambar 3. Hasil Kendall Tau Correlation untuk perbandingan antara *Big Five* dari web eksternal yang dimasukkan ke dalam proses *learning style* program dengan hasil *learning style* program menggunakan cara pertama menunjukkan nilai korelasi 0.29 sedangkan untuk cara kedua menunjukkan nilai korelasi sebesar 0.21. Untuk perbandingan antara hasil *learning style* dari web eksternal dengan hasil *learning style* program menggunakan cara pertama memiliki nilai korelasi 0.29 sedangkan untuk cara kedua memiliki nilai korelasi 0.21.

Untuk hasil responden kedua dapat dilihat pada Gambar 4. Hasil Kendall Tau Correlation untuk perbandingan antara *Big Five* dari web eksternal yang dimasukkan ke dalam proses *learning style* program dengan hasil *learning style* program menggunakan cara pertama menunjukkan nilai korelasi 0 sedangkan untuk cara kedua menunjukkan nilai korelasi sebesar -0.29.



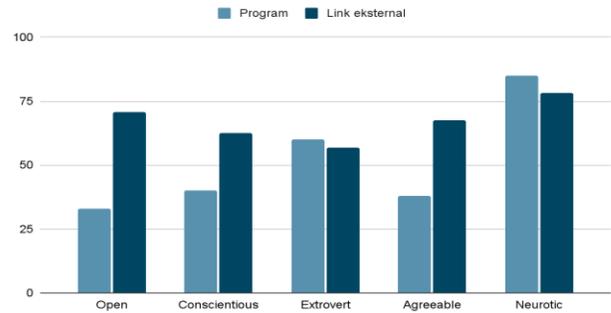
Gambar 3. Grafik Hasil Big Five Personality Responden Pertama

Untuk perbandingan antara hasil *learning style* dari web eksternal dengan hasil *learning style* program menggunakan cara pertama memiliki nilai korelasi -0.29 sedangkan untuk cara kedua memiliki nilai korelasi -0.21. Berikut adalah hasil survei kecocokan hasil responden kedua. Dari hasil kecocokan, responden kedua merasa bahwa hasil *Big Five* maupun *learning style* dari website eksternal lebih cocok daripada hasil program.

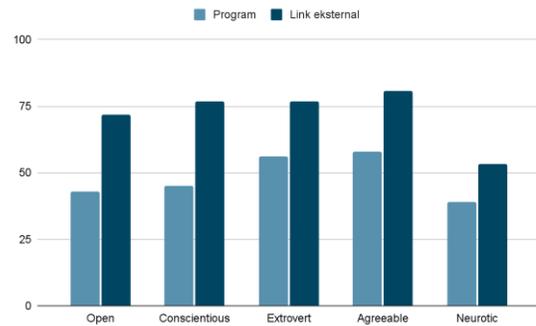
Untuk hasil responden kedua dapat dilihat pada Gambar 5. Hasil Kendall Tau Correlation untuk perbandingan antara *Big Five* dari web eksternal yang dimasukkan ke dalam proses *learning style* program dengan hasil *learning style* program menggunakan cara pertama menunjukkan nilai korelasi 0.78 sedangkan untuk cara kedua menunjukkan nilai korelasi sebesar 0.79. Untuk perbandingan antara hasil *learning style* dari web eksternal dengan hasil *learning style* program menggunakan cara pertama memiliki nilai korelasi 0.57 sedangkan untuk cara kedua memiliki nilai korelasi -0.29.

Untuk hasil responden ketiga dapat dilihat pada Gambar 6. Hasil Kendall Tau Correlation untuk perbandingan antara *Big Five* dari web eksternal yang dimasukkan ke dalam proses *learning style*

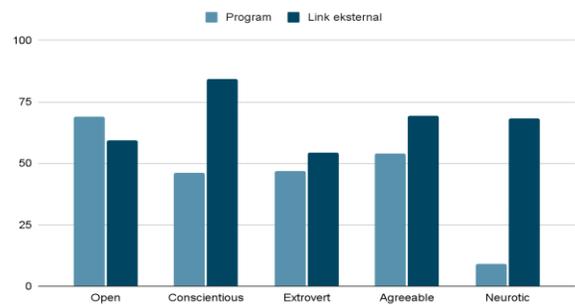
program dengan hasil *learning style* program menggunakan cara pertama menunjukkan nilai korelasi 0.29 sedangkan untuk cara kedua menunjukkan nilai korelasi 0.14. Perbandingan antara hasil *learning style* dari web eksternal dengan hasil *learning style* program menggunakan cara pertama memiliki nilai korelasi -0.07 dan cara kedua memiliki nilai korelasi 0.29.



Gambar 4. Grafik Hasil Big Five Personality Responden Kedua



Gambar 5. Hasil Big Five Personality Responden Ketiga

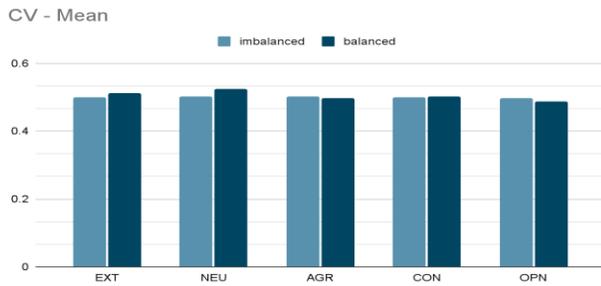


Gambar 6. Hasil Big Five Personality Responden Keempat

5.4 Pengujian Random Forest Dengan Data Twitter Yang Sudah Diseimbangkan Dengan Yang Belum Diseimbangkan

Data *balanced* yang digunakan, diproses seperti dengan data *imbalanced*. Dalam menentukan parameter yang paling tepat, digunakan juga *GridSearchCV* dan dua cara yang digunakan dalam

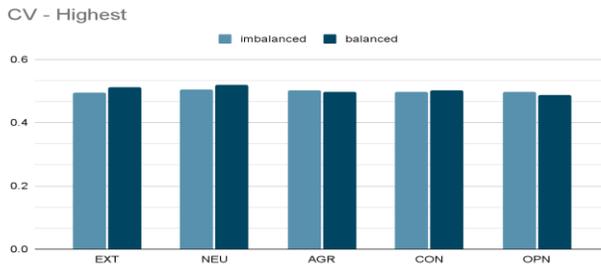
pengujian 5.1 juga dilakukan di pengujian ini. Hasil perbandingan dapat dilihat pada Gambar 7 dan Gambar 8.



Gambar 7. Perbandingan Data Imbalanced dan Balanced Dengan Cara Mean

5.5 Pengujian Learning Style

Untuk pengujian terhadap *learning style*, terdapat dua cara yang digunakan. Cara pertama dengan melakukan perkalian antara hasil *Big Five Personality* dengan nilai korelasi *Big Five Personality - learning style* yang dipakai kemudian dijumlahkan semua hasilnya dan dikalikan 0.11. Kemudian nilai tersebut di normalisasi dan dijumlahkan semua hasilnya setelah itu dikalikan 0.11.



Gambar 8. Perbandingan Data Imbalanced dan Balanced Dengan Cara Highest

Cara kedua dengan melakukan perkalian antara hasil *Big Five Personality* dengan nilai korelasi *Big Five Personality - learning style* yang memiliki nilai confidence dibawah 0.05 dan 0.01. Dari seluruh hasil *learning style* yang didapatkan dari keempat responden, masing-masing hasil tersebut dimasukkan ke dalam Kendall Tau untuk melihat bagaimana kecocokan dari *learning style* yang didapat dari program dengan hasil dari web eksternal. Hasil perbandingan dapat dilihat pada Tabel 3.

Tabel 3. Hasil Perbandingan Kendall Tau

	Eksperimen 1		Eksperimen 2	
	LS 1	LS 2	LS 1	LS 2
Responden 1	0.29	0.29	0.21	0.21
Responden 2	0	-0.29	-0.29	-0.21
Responden 3	0.78	0.57	0.79	-0.29
Responden 4	0.29	-0.07	0.14	0.29
Rata-rata	0.34	0.55	0.21	0

Eksperimen 1 merupakan perbandingan hasil *Big Five* dari web eksternal dimasukkan ke dalam *learning style* program

menggunakan cara pertama (LS 1) dan cara kedua (LS 2). Kemudian hasil tersebut dibandingkan dengan *learning style* dari web eksternal. Sedangkan eksperimen 2 merupakan perbandingan hasil *Big Five* dari hasil program dimasukkan ke dalam *learning style* program menggunakan cara pertama (LS 1) dan cara kedua (LS 2). Kemudian hasil tersebut dibandingkan dengan *learning style* dari web eksternal.

6. KESIMPULAN

Hasil akurasi Random Forest untuk dataset *balanced* dan *imbalanced* tidak terlalu signifikan, seperti *personality CON imbalanced* memiliki akurasi sebesar 0.499 sedangkan CON *balanced* memiliki akurasi sebesar 0.502 atau juga seperti *personality NEU imbalanced* memiliki akurasi sebesar 0.502 sedangkan NEU *balanced* memiliki akurasi sebesar 0.519. *Learning style* yang didapatkan dari program memiliki nilai kecocokan menurut korelasi Kendall Tau sebesar 0.21 jika dibandingkan dengan *learning style* responden. Parameter Random Forest seperti *n_estimator* dan *max_depth* lebih berpengaruh terhadap nilai akurasi sedangkan parameter *max_features* tidak terlalu berpengaruh terhadap nilai akurasi. Menurut survei kecocokan yang telah disebar keempat responden, responden lebih merasa web eksternal lebih cocok dengan nilai rata-rata 4.5 untuk *Big Five Personality* dan 4 untuk hasil *learning style*.

7. SARAN

Berikut adalah beberapa saran yang dapat dilakukan untuk mengembangkan penelitian lebih lanjut:

1. Mencoba menggunakan feature extraction yang berbeda, selain LIWC.
2. Menggunakan dataset yang jumlahnya lebih banyak dan lebih bervariasi.
3. Mencoba menambahkan feature lainnya seperti banyaknya jumlah followers untuk menentukan kepribadian, tidak hanya berdasarkan data posting

8. REFERENSI

- [1] Abidin, N.H., Akmal, M., Mohd, N.N., Nincarean, D., Yusoff, N., Karimah, H., & H, A. 2020. Improving Intelligent Personality Prediction using Myers-Briggs Type Indicator and Random Forest Classifier.
- [2] Akrami, N., Fernquist, J., Isbister, T., Kaati, L., & Pelzer, B. 2019. Automatic Extraction of Personality from Text: Challenges and Opportunities. 2019 IEEE International Conference on Big Data (Big Data). DOI= 10.1109/bigdata47090.2019.9005467
- [3] Cohen, A., & Baruth, O. 2017. Personality, learning, and satisfaction in fully online academic courses. *Computers in Human Behavior*, 72, 1–12. DOI= 10.1016/j.chb.2017.02.030
- [4] Djigić, G., Stojiljković, S., & Marković, A. 2016. Personality Traits and Learning Styles of Secondary School Students in Serbia. *Bulgarian Comparative Education Society.*, 14(1)
- [5] Felder, R. M., & Brent, R. 2005. Understanding student differences. *Journal of engineering education*, 94(1), 57-72.
- [6] Forsey, C. 2021. Twitter, Facebook, or Instagram? Which Platform(s) You Should Be On. *HubSpot Blog*. URI= <https://blog.hubspot.com/marketing/twitter-vs-facebook>.

- [7] Han, S., Huang, H., & Tang, Y. 2020. Knowledge of words: An interpretable approach for personality recognition from social media. *Knowledge-Based Systems*, 105550. DOI= 10.1016/j.knosys.2020.105550.
- [8] Hunt, D. E. 1979. Learning style and student needs: an introduction to conceptual level, student learning styles: diagnosing and prescribing programs. *Reston, VA: National Association of Secondary School*.
- [9] Huston, J. L., & Huston, T. L. 1995. How learning style and personality type can affect performance. *The Health Care Supervisor*, 13(4), 38-45.
- [10] How It Works? LIWC. URI= <https://liwc.wpengine.com/how-it-works/>
- [11] İlçin, N., Tomruk, M., Yeşilyaprak, S. S., Karadibak, D., & Savcı, S. 2018. The relationship between learning styles and academic performance in TURKISH physiotherapy students. *BMC Medical Education*, 18(1). DOI= 10.1186/s12909-018-1400-2.
- [12] Komarraju, M., Karau, S. J., Schmeck, R. R., & Avdic, A. 2011. The Big Five personality traits, learning styles, and academic achievement. *Personality and Individual Differences*, 51(4), 472-477. DOI= 10.1016/j.paid.2011.04.019
- [13] Köseoğlu, Y. 2016. To What Extent Can the Big Five and Learning Styles Predict Academic Achievement. *Journal of Education and Practice*, 7(30). ISSN: 2222-288X
- [14] Litzinger, T. A., Ha, L. S., Wise, J. C., & Felder, R. M. 2005. A Study of the Reliability and Validity of the Felder-Soloman Index of Learning Styles. *American Society for Engineering Education Annual Conference & Exposition*
- [15] Ong, V., Rahmanto, A. D. S., Williem & Suhartono. 2017. Exploring Personality Prediction from Text on Social Media: A Literature Review, *INTERNETWORKING Indones J*, 9, pp. 65-69. ISSN: 1942-9703
- [16] Siddiquei, N. L. & Khalid, R. 2018. The relationship between Personality Traits, Learning Styles and Academic Performance of E-Learners, 10, 249-263. DOI= : 10.5944/openpraxis.10.3.870
- [17] Tandera, T., Hendro, Suhartono, D., Wongso, R., & Prasetyo, Y. L. 2017. Personality Prediction System from Facebook Users. *Procedia Computer Science*, 116, 604-611. DOI= 10.1016/j.procs.2017.10.016
- [18] Testaccount93. 2019. *twitter_my_personality*. Kaggle. URI= <https://www.kaggle.com/testaccount93/twitter-my-personality>