

Machine Learning Prediksi Karakter Pengguna Hastag (#) Bahasa Generasi Milenial Di Sosial Media

Jonson Manurung, Hiskia Monel
Cibro, Siska Feronika Br. Sitepu

Info Artikel

Kata Kunci :

Hashtags, Instagram, Klasifikasi, K-Nearest Neighbors, Machine Learning

Keywords :

Hashtag, Instagram, Classification, K-Nearest Neighbors, Machine Learning

Abstrak

Setiap aktivitas pada internet (online) meninggalkan jejak digital yang dapat ditelusuri. Pengguna yang ekspresif bersosial media dan mempunyai kebiasaan menuangkan segala sesuatu pada Instagram kurang memperhatikan sebab akibat dari update status. Permasalahan yang dibahas dalam penelitian ini yaitu mendiskripsikan karakter user account Instagram sesuai hastag (tagar/#) bahasa milenial yang banyak dipergunakan seperti #thinking, #awesome dan seterusnya. Dengan machine learning, komputer dapat bekerja sendiri. Teknologi digital ini sudah lama diterapkan pada google search, search engine dan sosial media (facebook, twitter, instagram). Manfaat dari machine learning yaitu kemudahan memperoleh data digital user. Penelitian ini menghasilkan prediksi untuk classification menggunakan Algoritma K-Nearest Neighbors (KNN). Rumusan permasalahan dalam penelitian ini yaitu bagaimana mengolah data yang bersumber dari hastags (#) bahasa milenial berdasarkan the most popular hastags (#) on Instagram menggunakan machine learning dengan mengidentifikasi nama dalam teks ke dalam klasifikasi Connected, Creative dan Confident. Dari hasil perhitungan jarak terdekat dan kedekatan bertetangga diperoleh 10 hastag populer. Hasil klasifikasi diurutkan dari terkecil sampai terbesar diperoleh Hastag = #smart Post Count = 8928409. User dominan pada klasifikasi Creative..

Abstarct

Activity on the internet leaves a traceable digital trail. Users who are expressive of social media and have a habit of pouring everything on Instagram are more considerate of the cause and effect of status updates. The problem discussed in this study is to describe the character of the Instagram user account according to the hashtags (#) of the most widely used millennial language such as #awesome and so on. With machine learning, computers can work alone. This digital technology has long been applied to Google search, search engines and social media (Facebook, Twitter, Instagram). The benefits of machine learning are the ease of obtaining digital data from online users. The stages of the study consisted of the application of algorithms that produced predictions for classification using the K-Nearest Neighbors Algorithm. The formulation of the problem in this research is how to process data sourced from millennial language hashtags based on the most popular hashtags (#) on instagram using machine learning by identifying names in the text into Connected, Creative and Confident. From the results of the calculation of the closest distance and proximity of the neighboring obtained 10 popular hashtags. Creative Classifications become dominant type user.

1. PENDAHULUAN

Generasi milenial memiliki karakter unik berdasarkan wilayah dan kondisi sosial-ekonomi. Salah satu ciri utama generasi milenial ditandai oleh peningkatan penggunaan dan keakraban dengan komunikasi, media, dan teknologi digital. Dibesarkan oleh kemajuan teknologi, generasi milenial memiliki ciri-ciri kreatif, informatif, mempunyai passion dan produktif. Penggunaan bahasa milenial, muda maupun tua menunjukkan karakter kepribadian dari orang tersebut misalnya #richman, #hoerangkaya. Karakter dapat dibentuk dan dipengaruhi oleh orang-orang dan lingkungan tempat tinggal. Karakter juga dibentuk dari proses yang panjang dan lama. Sikap yang meyakini bahwa akan terlihat kekinian jika menggunakan istilah yang lagi trend (viral). Cara berkomunikasi mempengaruhi karakter seseorang. Permasalahan yang muncul individu (netizen) dituntut mengerti dan mengimplementasikannya dalam komunikasi sehari-hari. Ekspresi bahasa milenial mudah dilihat pada media sosial dengan tagar atau *hashtag*(#) *Instagram*, *Twitter*. Data bisa sama, akan tetapi algoritma dan pendekatan yang berbeda-beda untuk mendapatkan hasil yang optimal. Data terdiri dari data training dan data testing. Data training digunakan untuk melatih algoritma untuk mencari model yang cocok, sementara data testing akan dipakai untuk mengetahui performa model yang didapatkan pada tahapan testing. Dari model yang diperoleh dapat dilakukan prediksi. machine learning pada dasarnya adalah proses komputer untuk belajar dari data (learn from data). Data training diperoleh dari kumpulan hastag bahasa milenial yang dapat mempersentasikan karakter account ig. Hastags bahasa milenial dipergunakan pada sosial media untuk mempertegas penyampaian informasi dan mengidentifikasi karakter user account media sosial Instagram [1]. Hastags bahasa milenial akan viral bila dipergunakan berulang-ulang oleh user account yang lain, ini akan menambah kepopuleran dan jumlah pengguna. Model yang dihasilkan dapat melakukan klasifikasi atau prediksi kedepannya menggunakan Machine Learning. Terdapat ribuan bahkan lebih bahasa milenial untuk itu terlebih dahulu mengidentifikasi hastags bahasa milenial yang sedang viral. Dilajut mengolah data yang bersumber dari hastags (#) bahasa milenial berdasarkan the most popular hastags (#) on Instagram menggunakan Machine Learning [2].

K-nearest neighbors melakukan klasifikasi dengan proyeksi data pembelajaran pada ruang berdimensi banyak. Ruang ini dibagi menjadi bagian-bagian yang merepresentasikan kriteria data pembelajaran. Setiap data pembelajaran direpresentasikan menjadi titik-titik c pada ruang dimensi banyak. Untuk menggunakan algoritma k nearest neighbors, perlu ditentukan banyaknya k tetangga terdekat yang digunakan untuk melakukan klasifikasi data baru. Penentuan nilai k dipertimbangkan berdasarkan banyaknya data yang ada dan ukuran dimensi yang dibentuk oleh data. Semakin banyak data yang ada, angka k yang dipilih sebaiknya semakin rendah. semakin besar ukuran dimensi data, angka k yang dipilih sebaiknya semakin tinggi. Data baru yang diklasifikasi selanjutnya diproyeksikan pada ruang dimensi banyak yang telah memuat titik-titik c data pembelajaran. Proses klasifikasi dilakukan dengan mencari titik c terdekat dari c -baru (Nearest Neighbor). Teknik pencarian tetangga terdekat yang umum dilakukan dengan menggunakan formula jarak Euclidean. Dalam metode KNN penggunaan atribut kata yang hadir pada suatu dokumen menjadi dasar klasifikasi untuk penggolongan atau klasifikasi teks.

Penelitian [14] berjudul Pengolahan Data Menggunakan Machine Learning menjelaskan di dalam data set terdapat 10 iklan berbeda yang akan ditunjukkan kepada 10000 orang, kemudian algoritma akan menentukan 1 dari 10 iklan yang diketahui sebagai iklan yang paling tepat untuk 10000 orang. Penelitian [1] berjudul Identifikasi Hoax pada Media Sosial dengan Pendekatan Machine Learning memaparkan metode klasifikasi tanpa fitur menunjukkan performa akurasi yang baik dalam mengidentifikasi konten media sosial [3].

2. METODE PENELITIAN

Machine Learning merupakan kecerdasan buatan (Artificial Intelligence) yang mempelajari bagaimana cara mengembangkan suatu sistem, membuat data, dan membuat algoritma-algoritma yang bisa melakukan tugas-tugas dengan sendiri tanpa ada arahan dari pengguna dan memungkinkan programmer bisa belajar [4]. Machine learning adalah metode analisis yang membantu menangani data besar dengan cara mengembangkan algoritma komputer. Konsep machine learning yaitu mampu meningkatkan kecerdasan yang meliputi kemampuan dari suatu individu untuk belajar tanpa terkecuali pada sebuah mesin. Produktivitas manus akan meningkat jika suatu mesin mampu untuk belajar. Dan mesin tersebut akan berbeda dengan mesin lain, karena mempunyai kemampuan yang tidak dimiliki oleh mesin lain. Secara fundamental cara kerja machine learning adalah belajar seperti manus dengan menggunakan contoh-contoh dan setelah itu barulah dapat menjawab suatu pertanyaan terkait [5]. Proses belajar ini menggunakan data yang disebut train dataset. Dari data tersebut, komputer akan melakukan proses belajar (training) untuk menghasilkan suatu model. Proses belajar ini menggunakan algoritma machine learning sebagai penerapan teknik statistika. Model inilah yang menghasilkan informasi, kemudn dapat dijadikan pengetahuan untuk memecahkan suatu permasalahan sebagai proses *input-output* [6].

2.1 Metode Machine Learning

Metode Algoritma Machine Learning:

1. *Supervised machine learning algorithms*

Supervised machine learning menerapkan informasi yang telah ada pada data dengan memberikan label tertentu, misalnya data yang telah diklasifikasikan sebelumnya (terarah). Algoritma ini mampu memberikan target terhadap output yang dilakukan dengan membandingkan pengalaman belajar di masa lalu.

2. *Unsupervised machine learning algorithms*

Unsupervised machine learning digunakan pada data yang tidak mempunyai informasi yang dapat diterapkan secara langsung (tidak terarah). Algoritma ini diharapkan mampu menemukan struktur tersembunyi pada data yang tidak berlabel.

2. *Semi-supervised machine learning algorithms*

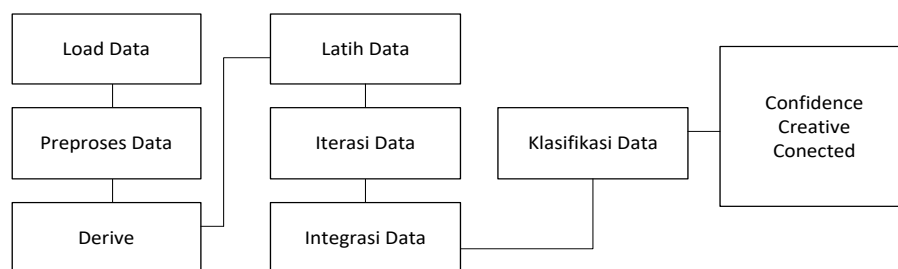
Semi-supervised machine learning digunakan untuk melakukan pembelajaran data berlabel dan tanpa label. Sistem yang menggunakan metode ini dapat meningkatkan efisiensi output yang dihasilkan.

4. *Reinforcement machine learning algorithms*

Reinforcement machine learning mempunyai kemampuan untuk berinteraksi dengan proses belajar yang dilakukan, algoritma ini akan memberikan poin (*reward*) saat model yang diberikan semakin baik atau mengurangi poin (*error*) saat model yang dihasilkan semakin buruk. Salah satu penerapannya adalah pada mesin pencari [7] [8].

2.1.1 Tahapan Penelitian

Pembelajaran mesin mengenali secara otomatis pola kompleks dan membuat keputusan cerdas berdasarkan data. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Uraian tahapan penelitian sebagai berikut :

1. Akses dan load data.

Preprocessing merupakan tahapan proses dalam mempersiapkan data mentah yang dapat diolah lebih lanjut. Praproses data dilakukan dengan cara mengubah data yang tidak sesuai, menjadi bentuk yang lebih mudah diproses oleh sistem.

2. Proses Data

Mengubah data mentah menjadi informasi. Bagi data menjadi dua set. Sebagian data untuk menguji (set tes) dan menggunakan sisanya (set pelatihan) untuk membangun model.

3. Derive menggunakan fitur data yang diproses sebelumnya. Hitung nilai turunan yang meningkat.

4. Latih Data

Melatih data menggunakan model fitur yang diperoleh pada langkah membangun model.

5. Iterasi Data

Iterasi untuk menemukan model terbaik. KNN, tetangga terdekat sebuah algoritma sederhana yang menyimpan semua data pelatihan, membandingkan poin baru ke melatih data.

6. Mengintegrasikan model yang paling terlatih

menjadi sistem produksi. Meningkatkan model dapat mengambil dua arah yang berbeda: memodelkan lebih sederhana atau menambah kompleksitas.

7. Klasifikasi menggunakan tetangga terdekat

Model klasifikasi [9] [10]:

- 1) Membangun klasifikasi KNN
- 2) Periksa kualitas klasifikasi KNN
- 3) Prediksi klasifikasi menggunakan KNN Classifier

4) Tentukan KNN Classifier

K-nearest neighbors atau KNN adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran (train data sets), yang diambil dari k tetangga terdekatnya (nearest neighbors) [11]. Dengan k merupakan banyaknya tetangga terdekat. K-nearest neighbors melakukan klasifikasi dengan proyeksi data pembelajaran pada ruang berdimensi banyak. Untuk menggunakan algoritma k nearest neighbors, perlu ditentukan banyaknya k tetangga terdekat yang digunakan untuk melakukan klasifikasi data baru. Banyaknya k, sebaiknya merupakan angka ganjil, misalnya k = 1, 2, 3, dan seterusnya. Penentuan nilai k dipertimbangkan berdasarkan banyaknya data yang ada dan ukuran. Jarak Euclidean adalah formula untuk mencari jarak antara 2 titik dalam ruang dua dimensi. Formula yang digunakan dalam algoritma KNN, Perhitungan metode K-Nearest Neighbor, data dibagi atas data latih dan data uji. Teknik perhitungan untuk mencari jarak menggunakan Euclidean, digunakan rumus Euclidean [12] [13], persamaan 1.

$$d(x_i - x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \dots\dots\dots(1)$$

$(x_i - x_j)$ = Jarak Euclidean (Euclidean Distance). $(x_i), (x_j)$ = record ke-i , record ke-j ; a_r = data ke-r ; i,j : 1,2,3,...n ; n : dimensi objek

3. HASIL DAN PEMBAHASAN

Generasi milenial menggunakan bahasa kekinian di kehidupan sehari-hari atau saat beraktivitas di media sosial Instagram (#), dapat dilihat pada Tabel 1.

Tabel 1. Hashtag Populer Pada Instagram 2019

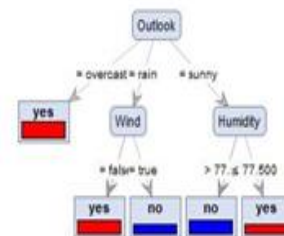
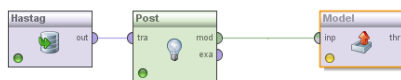
Ranking of the 10 More Used Hashtags	Hashtags is Constantly Growing
#Love has been used in 1,597,832,456 posts. #Instagood, in second place, has appeared in 988,131,008 publications. #Photooftheday: Over 490 million publications included this hashtag. #Fashion has been used in 475,992,147 publications on Instagram so far. #Beautiful is the fifth most popular hashtag with a total of 456,547,737 tags. #Happy 422,116,954 publications. #Like4like 410,451,976 publications. #Picoftoday falls behind with 373,976,698. #TBT with 489.000.000 impressions #Art has been used in just over 334 million photos or videos uploaded on Instagram.	#instagram #followme #style #follow #instadaily #travel #life #cute #fitness #nature #beauty #girl #fun #photo #amazing #likeforlike #instalike #Selfie #smile #me #lifestyle #model #follow4follow #music #friends #motivation #like #food #inspiration #Repost #summer #design #makeup #TBT #followforfollow #ootd #Family #141 #cool #igers #TagsForLikes #hair #instamood #sun #vsco #fit #beach #photographer #gym #artist #girls #vscocam #autumn #pretty #luxury #instapic #black #sunset #funny #sky #blogger #hot #healthy #work #bestoftheday #workout #f4f#nofilter #london #goals #blackandwhite #blue #swag #health #party #night #landscape #nyc#happiness #pink #lol #foodporn #newyork #fitfam #awesome #fashionblogger #Halloween #Home #fall #paris

Data baru yang diklasifikasi selanjutnya diproyeksikan pada ruang dimensi banyak yang telah memuat titik-titik c data pembelajaran. Proses klasifikasi dilakukan dengan mencari titik c terdekat dari c-baru (nearest neighbor). Teknik pencarian tetangga terdekat yang umum dilakukan dengan menggunakan formula *jarak euclidean*. Beberapa aplikasi disediakan sebagai mesin pencari. Machine learning menjadi mesin tersendiri bekerja sendiri sebagai searching text mining. Mesin Pencarian hashtag untuk mengenal trending yang menunjukkan karakter pengguna bahasa generasi milenial melalui link <https://www.tagsfinder.com/id-id/>; <https://www.digitalinformationworld.com/15/07/infographic-how-hashtags-impact-instagram-engagement.html?m=1>.

Pengolahan data metode K-Nearest Neighbor menggunakan data sebanyak 100 hashtag. Diambil hashtag trending paling mendekati pengklasifikasian karakter antara lain #starting: 514.914 #handmade 196.296542 #smart: 8.928.409 #share: 24.435.48 #awesome: 95.534.084 #beyourself: 13.527.498 #opendmind: 811.578 #idea: 6.012.997 #makesomething: 362.493 #winner 13.495.340. Pengklasifikasian hashtag menurut karakter dibagi 3 yaitu Connected, Creative dan Confidence dengan variabel Hashtag dan Jumlah Post. Pemodelan trending hashtag menggunakan Ripedminer 5.0. Operator membaca file input format .xls.

Tabel 2. Proses Load Data

Load Data				Pemodelan Klasifikasi		Decision Tree
Row No.	Hastag	Jumlah Post	Klasifikasi	Hashtag	Post	Model
1	#starting	514914.0	Confidence	Hashtag	Post	Model
2	#handmade	1.96296542E7	Creative			
3	#smart	8928409.0	Connected			
4	#share	2443548.0	Connected			
5	#awesome	95.534.084	Confidence			
6	#beyourself	1.3527498E7	Confidence			
7	#opendmnc	811578.0	Creative			
8	#idea	6012997.0	Creative			
9	#makesome	362493.0	Creative			
10	#winner	1.349534E7	Confidence			



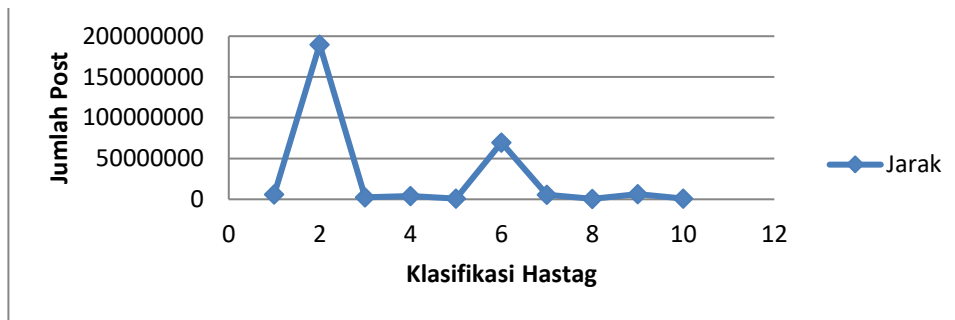
Tabel 2 menunjukkan 10 dari 100 data testing sebagai pengujian hasil model terbaik yang dihasilkan K-NN. Dari data ini akan dihasilkan kesimpulan sebuah data akan masuk kedalam pengklasifikasian karakter. Proses perhitungan persamaan Eulidean Distance dilakukan pada setiap data lama sehingga nantinya akan menghasilkan nilai jarak sesuai dengan jumlah data lama. Pencarian tetangga terdekat (k) dari Row 9 Hastag: #makesomething Jumlah Post = 362493 klasifikasi = Creative.

Sebelum melakukan proses perhitungan KNN disiapkan data kelas/ label, data atribut, data lama (training), dan data baru (testing). Jangkauan dan atribut penentu yang digunakan untuk menentukan kelas pada klasifikasi. Input data sesuai dengan atribut (hastag, jumlah post). Hitung jarak nilai atribut testing ke setiap atribut training menggunakan persamaan Euclidean Distance. Banyaknya k tetangga terdekat yang digunakan untuk melakukan klasifikasi data. baru. k, sebaiknya merupakan angka ganjil, misalnya k = 1, 2, 3, dan seterusnya. Penentuan nilai k dipertimbangkan berdasarkan banyaknya data yang ada dan ukuran dimensi yang dibentuk oleh data. Semakin banyak data yang ada, angka k yang dipilih sebaiknya semakin rendah. Semakin besar ukuran dimensi data, angka k yang dipilih sebaiknya semakin tinggi. Sorting hasil jarak berdasarkan dari nilai terendah. Hitung Jumlah Nilai yang paling banyak, prediksi selanjutnya hitung akurasi. Hasil perhitungan jarak, pada Tabel 3.

Tabel 3. Hasil Perhitungan Jarak

Hastag	Jumlah Post	Klasifikasi	Jarak	Urutan
#smart	8928409	Connected	571169	3
#idea	6012997	Creative	691174	8
#starting	514914	Confidence	889918	1
#beyourself	13527498	Confidence	2344243	6
#makesomething	362493	Creative	4140618	9
#share	2443548	Connected	5772588	4
#handmade	196296542	Creative	6069252	2
#winner	13495340	Confidence	6221673	10
#opendmind	811578	Creative	69433332	7
#awesome	95534084	Confidence	189713276	5

Dari hasil training, diperoleh data uji baru #thinking karakter klasifikasi Connected Jumlah Post 6584166. Klasifikasi objek berdasarkan atribut dan training sample. Clasifier tidak menggunakan apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik query, akan ditemukan sejumlah k obyek atau (titik training) yang paling dekat dengan titik query. Klasifikasi menggunakan pemodelan data Machine Learning dari k obyek. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari query instance yang baru. Untuk pemilihan k lainnya klasifikasi ditentukan frekuensi terbanyak. Pada penelitian ini, metode K-NN mencapai hasil akurasi yang baik dan sesuai dengan perhitungan yang diterapkan.



Gambar 2. Grafik Klasifikasi User Dengan K-Nearest Neighbor

4. KESIMPULAN

Kesimpulan dari hasil dan pembahasan :

1. Algoritma K-Nearest Neighbor paling banyak digunakan untuk klasifikasi dengan mengelompokkan data, kemiripan atau kedekatan yang ada pada training data. Semakin banyak data yang training maka akan semakin besar akurasi yang dapat dihasilkan dari testing.
2. Load data menghasilkan data yang telah dibersihkan (celansing data) selanjutnya dipisahkan, dikelompokkan, dengan pendekatan Machine learning.
3. Penerapan K-Nearest Neighbor memberi kemudahan dalam menentukan klasifikasi pengguna hashtag user Instagram dalam memprediksi tipe karakter user dengan akurasi yang baik.

UCAPAN TERIMA KASIH

Ucapan Terima Kasih kepada STMIK Pelita Nusantara dengan Surat Penugasan Dalam Rangka Pelaksanaan Program Penelitian/ Pengabdian Tahun Anggaran 2019 Nomor : 011/PN/10/05/PM/2019..

DAFTAR PUSTAKA

- [1] P. K. L. Utama, "Identifikasi Hoax pada Media Sosial dengan Pendekatan Machine Learning", Widya Duta Vol. 13, No. 1 2018, hal 69-76.
- [2] Purnomo, M, H, Sumpeno, S, Setiawan, E, I, Purwitasari, D, "Biomedical Engineering Research in the Social Network Analysis Era: Stance Classification for Analysis of Hoax Medical News in Social Media", Procedia Computer Science 116 2017, pp. 3-9.
- [3] R, R, Sani, J. Zeniarja, A. Luthfiarta, "Penerapan Algoritma K-Nearest Neighbor pada Information Retrieval dalam Penentuan Topik Referensi Tugas Akhir", Journal of Applied Intelligent System, Vol. 1, No. 2, Juni 2016: 123 – 133.
- [4] S. M. Hutabarat, Anita Sindar, "Data Mining Penjualan Suku Cadang Sepeda Motor Menggunakan Algoritma K-Means", Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI) 2 (2), 126-132.
- [5] M. Rivki, A. M. Bachtiar, "Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia", Jurnal Sistem Informasi (Journal of Information Systems), 2017, 31-37.
- [6] D. Simanjuntak, A. S. RM. Sinaga, "Sistem Pakar Deteksi Gizi Buruk Balita Dengan Metode Naïve Bayes Classifier", Jurnal Inkofar * Volume 1 No. 2 Desember 2019, hl: 54-60.
- [7] Chairani, Widyawan, S. S. Kusumawardani, "Machine Learning Untuk Estimasi Posisi Objek Berbasis RSS Fingerprint Menggunakan IEEE 802.11g Pada Lantai 3 Gedung JTETI UGM", Jurnal Infotel Vol. 7 No. 1 Mei 2015.
- [8] D. A. Ambarsari, R. N. Nurfalah, S. J. Kuryanti, "Penerapan Deep Learning dalam Pendeteksian Autism Toddler", Infotekjar :Jurnal Nasional Informatikadanteknologijaringan- Vol. 4 No. 1, 2019.
- [9] F. Thabtah, "Autism spectrum disorder screening: Machine learning adaptation and DSM-5 fulfillment," ACM Int. Conf. Proceeding Ser., vol. Part F129311, 2017, pp. 1-6.
- [10] Mustakim, G. Oktaviani, "Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa", Jurnal Sains Teknologi dan Industri, Vol. 13, No.2, Juni 2016, pp.195 – 202.
- [11] A. S. RM. Sinaga, R. N. Zendrato, "Optimasi Penugasan Pegawai Menggunakan Metode Hungarian", Journal

- of Innovation Information Technology and Application (JINITA), Volume 1 Issue 01 Pages 16-24 Tahun 2019.
- [12] Razzaghi, T., Roderick, O., Safro, I., Marko, N, Multilevel weighted support vector machine for classification on healthcare data with missing values. PloS ONE, 11(5), e0155,119.P. K. L. Utama. 2018. Identifikasi Hoax pada Media Sosial dengan Pendekatan Machine Learning”, Widya Duta Vol. 13, No. 1. hal 69-76.
- [13]. Ramadhan, R. S. Junta, Z., Ardytha, L. 2016. Penerapan Algoritma K-Nearest Neighbor pada Information Retrieval dalam Penentuan Topik Referensi Tugas Akhir. Journal of Applied Intelligent System, Vol. 1, No. 2, 123 – 133.
- [14] Rauhan, Ahmad. 2019. Pengolahan Data Menggunakan Machine Learning. [online] Available : <https://library.universitaspertamina.ac.id/xmlui/bitstream/handle/123456789/162/Jurnal%20Ilmiah.pdf?sequence=1&isAllowed=y>