

METODE PEMBELAJARAN MESIN UNTUK MENILAI DATA DEPRESI DAN KESEHATAN MENTAL

MACHINE LEARNING METHODS FOR ASSESSING DEPRESSION AND MENTAL HEALTH DATA

Maritza Anastasia¹, Vivia Surya Maulivia², Suharjito^{3*}

^{1,2,3}Magister Teknik Industri, Fakultas Teknik, Universitas Bina Nusantara

maritza.anastasia@binus.ac.id¹, vivia.maulivia@binus.ac.id², suharjito@binus.ac.id³

ABSTRACT

One mental wellbeing clutter that frequently happens in people is misery. Recognizing misery early is critical for people. In any case, in truth, conducting early misery screening still has a few disadvantages. In case it proceeds to be disregarded, this could have an impact on person wellbeing. Subsequently, there's a require for other strategies that can speak to the level of discouragement in people. This inquire about points to decide the portrayal of workers' recognitions of the psychosocial components of work setting and substance and to decide the indications of work stretch they complain around. The subordinate variable is indications of work push (physical, mental, cognitive and behavioral), whereas the autonomous variable is person characteristic components (age, sexual orientation, sort of work and length of benefit). The classification method employments different classification models, to be specific the K-Nearest Neighbor (KNN), Choice Tree, Arbitrary Woodland Tree, Naïve Bayes, Back Vector Machine (SVM) and AdaBoost calculation models. The comes about of the investigation based on the comes about of testing different calculation models can be concluded that the leading calculation demonstrate for measuring specialist pay is utilizing K-Nearest Neighbor (KNN). Variable investigation based on Occupation and Sexual orientation factors predicts the level of discouragement and mental wellbeing filled by members matured 25-30 a long time.

Keywords: Depression and Mental Health, Machine Learning, Algorithm Method Comparison.

ABSTRAK

Salah satu gangguan kesejahteraan mental yang sering terjadi pada manusia adalah depresi. Mengenali depresi secara dini penting bagi individu. Namun, pada kenyataannya, melakukan skrining depresi secara dini masih memiliki beberapa kelemahan. Jika terus diabaikan, hal ini dapat berdampak pada kesejahteraan individu. Oleh karena itu, diperlukan metode lain yang dapat menggambarkan tingkat depresi pada individu. Penelitian ini bertujuan untuk menentukan gambaran persepsi pekerja tentang komponen psikososial dari lingkungan dan isi kerja serta menentukan gejala stres kerja yang mereka keluhkan. Variabel dependen adalah gejala stres kerja (fisik, mental, kognitif, dan perilaku), sedangkan variabel independen adalah karakteristik individu (usia, jenis kelamin, jenis pekerjaan, dan lama bekerja). Metode klasifikasi menggunakan berbagai model klasifikasi, yaitu K-Nearest Neighbor (KNN), Decision Tree, Random Forest Tree, Naïve Bayes, Support Vector Machine (SVM), dan model penghitungan AdaBoost. Hasil penelitian berdasarkan pengujian berbagai model penghitungan dapat disimpulkan bahwa model penghitungan terbaik untuk mengukur gaji pekerja adalah menggunakan K-Nearest Neighbor (KNN). Analisis variabel berdasarkan faktor Pekerjaan dan Jenis Kelamin memprediksi tingkat depresi dan kesejahteraan mental yang dirasakan oleh anggota berusia 25-30 tahun.

Kata Kunci: Depresi dan Kesehatan Mental, Pembelajaran Mesin, Perbandingan Metode Algoritma.

PENDAHULUAN

Kesejahteraan mental adalah bagian penting dari kesehatan. Kondisi mental yang sehat akan memungkinkan individu untuk mencapai potensi mereka, mengatasi stres dan masalah dalam hidup, menjadi produktif di tempat kerja, dan berkontribusi pada masyarakat (WHO, 2020) [1]. Gaya hidup modern memiliki dampak mental pada pikiran individu,

menyebabkan stres emosional dan depresi. Satu miliar orang menderita gangguan mental dan lebih dari 300 juta orang menderita depresi di seluruh dunia.

Depresi mendominasi pemikiran bunuh diri pada seseorang. Sekitar 800.000 orang melakukan bunuh diri setiap tahunnya. Oleh karena itu, respons komprehensif diperlukan untuk mengatasi beban masalah kesehatan mental. Depresi

adalah gangguan mental umum yang memengaruhi pemikiran dan perkembangan mental seseorang.

Menurut WHO, Organisasi Kesehatan Dunia, masalah yang sering terjadi pada masa remaja dan awal dewasa terkait dengan masalah kesehatan mental, salah satunya adalah depresi. Depresi merupakan penyebab utama beban penyakit di kalangan individu pada masa awal kehidupan (WHO, 2016). Menurut WHO, depresi merupakan penyebab utama penyakit dan ketidakmampuan pada kaum muda, dan bunuh diri merupakan penyebab kematian ketiga terbesar (WHO, 2014).

Menurut Pieper dan Uden (2006), kesehatan mental adalah kondisi di mana seseorang tidak merasa bersalah tentang dirinya sendiri, memiliki pandangan yang positif tentang dirinya sendiri, dan mampu menerima kelemahan dan kekurangan pribadinya. Kepuasan hidup, kepuasan hidup sosial, dan kebahagiaan hidup. Liga Kesehatan Mental Dunia menyatakan bahwa kesehatan mental adalah kondisi yang memungkinkan perkembangan fisik, mental, dan emosional yang optimal dalam harmoni dengan orang lain. Masyarakat yang mentalnya sehat adalah masyarakat yang memungkinkan anggotanya untuk berkembang sesuai dengan kemampuan mereka.

Pentingnya mengenali depresi secara dini diilustrasikan oleh peningkatan frekuensi bunuh diri di berbagai negara yang disebabkan oleh kurangnya pencegahan atau pengobatan dini dalam kasus depresi. Organisasi Kesehatan Dunia (WHO) menunjukkan bahwa lebih dari 264 juta orang dari semua usia menderita depresi di seluruh dunia. Sebuah komite dari Institut Kedokteran tentang Pencegahan Gangguan Mental telah mengidentifikasi depresi sebagai gangguan yang paling dapat dicegah. Oleh karena itu, sangat penting untuk mengidentifikasi individu yang menderita depresi secara dini untuk mengurangi dampaknya terhadap kesehatan masyarakat. Selain itu,

deteksi dini dapat mengurangi risiko kewirausahaan.

Pada kenyataannya, skrining depresi dini masih memiliki beberapa kelemahan. Hal ini disebabkan oleh ketidaktahuan masyarakat akan depresi dan kelalaian terhadapnya. Jika diabaikan secara terus-menerus, kesehatan masyarakat dapat terpengaruh. Oleh karena itu, teknik tambahan diperlukan untuk mengukur tingkat depresi seseorang.

Karya Terkait

Dengan menggunakan metode pembelajaran mesin seperti Pohon Random Forest (RFT), Mesin Vector Dukungan (SVM), dan Jaringan Neural Konvolusi (CNN), beberapa peneliti telah mencoba untuk memprediksi kecemasan dan kesedihan untuk kemudian mengumpulkan dan mengklasifikasikan data pos blog. Banyak pendekatan, termasuk pemodelan topik, Bag-of-Words (BOW), dan Term Frequency – Inverse Document Frequency (TF-IDF), telah digunakan untuk pengkodean teks. Selain itu, eksperimen pemodelan telah dilakukan menggunakan pemrograman Python. CNN menghasilkan hasil terbaik di antara semua klasifier [8], dengan skor akurasi dan recall sebesar 78% dan 0,72, secara berturut-turut. Menurut penelitian Eri Mardiani, "Perbandingan Metode kNN, Naive Bayes, Decision Tree, Ensemble, dan Regresi untuk Prediksi Pendapatan," Keputusan Naive Bayes adalah model algoritma terbaik untuk memperkirakan pendapatan pekerja.

Banyak algoritma pembelajaran mesin telah dikembangkan, termasuk RFT, SVM, Catboost, Naive Bayes, dan Regresi Logistik. telah mengajukan aplikasi untuk kategorisasi [9]. Dalam studi ini, 470 pelaut berpartisipasi dalam proses wawancara di mana mereka diberikan informasi tentang pekerjaan, sosiodemografi, dan kesehatan mereka. Informasi ini dikumpulkan melalui 16 karakteristik: usia, status pekerjaan, BMI, masa kerja, tipe keluarga, status

perkawinan, profil pekerjaan, peringkat dalam organisasi, jenis kapal yang diposting, dan variabel dummy untuk kualifikasi akademik dan status perkawinan.

Oleh karena itu, para peneliti menemukan bahwa Catboost menghasilkan akurasi dan presisi terbesar (82,6% dan 84,1%, secara berturut-turut) dari semua klasifier. Sau et al. (2017) secara manual mengumpulkan informasi tentang 630 orang tua, 520 di antaranya membutuhkan perawatan khusus, dari Kolese Kedokteran dan Rumah Sakit Kolkata, Bengal Barat. Setelah implementasi berbagai teknik klasifikasi, termasuk regresi logistik, multiple layer perceptrons, Naïve Bayes, hutan acak, pohon acak, J48, optimisasi berurutan acak, subspace acak, dan K-bintang, teramati bahwa hutan acak menghasilkan tingkat akurasi tertinggi, masing-masing mencapai 89% dan 91%, antara dua set data yang terdiri dari 110 dan 520 individu. Alat WEKA digunakan dalam makalah "Memprediksi kecemasan dan depresi pada pasien lanjut usia menggunakan pembelajaran mesin" untuk pemilihan fitur dan klasifikasi.

Saat ini, media sosial dengan cepat menjadi alat penilaian kesehatan untuk memprediksi berbagai jenis penyakit. Saha et al. [10] Memilih topik dan karakteristik psikolinguistik muncul dalam pos di situs LiveJournal. Kemudian dimasukkan ke dalam sistem pemodelan bersama, untuk mengklasifikasikan masalah mental yang terjadi dalam komunitas online yang ingin tahu tentang depresi. Sistem pemodelan bersama yang diusulkan melampaui pembelajaran tugas tunggal (STL) dan pembelajaran multi-tugas (MLT) yang ada, dan penelitian menunjukkan bahwa ada diskusi di komunitas online tentang lebih dari sekadar perasaan sedih. Reece et al. [11] berfokus pada indikator depresi dan Gangguan Stres Pasca Trauma (PTSD) di antara pengguna Twitter. Hidden Markov Model (HMM) digunakan untuk mengidentifikasi kemungkinan terjadinya

PTSD. Dari total data set, 31,4% dan 24% teramati memiliki depresi dan PTSD, secara berturut-turut. Braithwaite et al. [12] mengumpulkan tweet dari 135 anggota yang dipilih dari Amazon Mechanical Turk (MTurk) dan menggunakan pohon keputusan klasifikasi untuk mengukur risiko bunuh diri. Tingkat akurasi untuk memprediksi tingkat bunuh diri diketahui mencapai 92%.

Du et al. [13] mengekstraksi data streaming dari Twitter dan menggunakan pemicu stres psikiatri untuk menjelaskan tweet yang dianggap memiliki kecenderungan bunuh diri. Convolution Neural Network (CNN) mengungguli Mesin Vector Dukungan (SVM) dan pohon tambahan (ET) dll. dengan akurasi 78% dalam mengidentifikasi tweet dengan kecenderungan bunuh diri. Pendekatan audio-teks juga dapat digunakan untuk menunjukkan depresi, di mana para peneliti mengumpulkan data tentang orang dengan depresi. Model jaringan saraf memori jangka pendek digunakan untuk mendeteksi depresi dalam studi "Mendeteksi Depresi dengan Pemodelan Urutan Audio/Teks" [14], yang menemukan bahwa model bebas konteks memberikan hasil terbaik untuk audio (bobot, berurutan, dan multi-model).

Kecemasan dan gangguan suasana hati diidentifikasi dalam studi "Menggunakan Pembelajaran Mesin untuk Menggolongkan Subtipe Berbasis Sirkuit dalam Gangguan Suasana Hati dan Kecemasan" [15] dengan memfilter ekspresi wajah pasien dan menerapkan metode validasi silang, dan hasil yang lebih akurat ditemukan yang dikonfirmasi oleh berbagai ukuran statistik yang berbeda. Klasifikasi ketidakseimbangan diterapkan dalam [16] dan teknik pembelajaran mesin berkelompok dibahas dalam penelitian "Teknik Ensemble dalam Pembelajaran Mesin." Dalam Workshop Internasional tentang Sistem Klasifikasi Berbeda Springer" [17]. Para peneliti yang berbeda telah menerapkan berbagai algoritma pembelajaran mesin untuk

prediksi gangguan mental, dan kinerja algoritma yang berbeda ternyata bervariasi, tergantung pada situasi; Tidak ada algoritma yang tetap telah ditentukan sebagai yang paling sesuai dalam semua kasus. Oleh karena itu, dalam penelitian ini, semua algoritma pembelajaran mesin diterapkan untuk mengenali gejala kecemasan, kesedihan, dan stres.

METODE

The purpose of this study is to identify the Gejala stres kerja yang dilaporkan oleh karyawan, serta untuk menjelaskan bagaimana mereka mempersepsikan aspek psikososial dari pengaturan dan konten pekerjaan. Sebanyak 825 peserta dari berbagai latar belakang—termasuk remaja dari Bangladesh, mahasiswa, ibu rumah tangga, profesional dari perusahaan dan firma, dan orang lain—berpartisipasi dalam penelitian ini. Pria dan wanita antara usia 16 dan 30 tahun diundang untuk mengisi kuesioner. Enam teknik pembelajaran mesin kemudian digunakan untuk mengklasifikasikannya: K-Nearest Neighbor (KNN), Decision Tree, Random Forest Tree, Naïve Bayes, Support Vector Machine (SVM), dan AdaBoost. Untuk penelitian ini, data primer dan sekunder dikumpulkan. Sumber data primer adalah kumpulan data analitis untuk kesehatan mental dan depresi yang diperoleh dari situs web Kaggle. <https://www.kaggle.com/datasets/shashwat/work/depression-and-mental-health-data-analysis>.

Media online dan sumber lainnya menyediakan metode pengumpulan data sekunder untuk penelitian ini. Peneliti akan menggunakan tinjauan literatur untuk menjelaskan dan mengevaluasi data yang dikumpulkan untuk penelitian ini (Dachi, 2023).

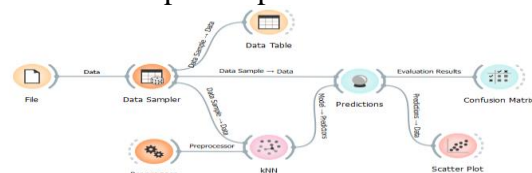


Gambar 1. Tahapan Penelitian

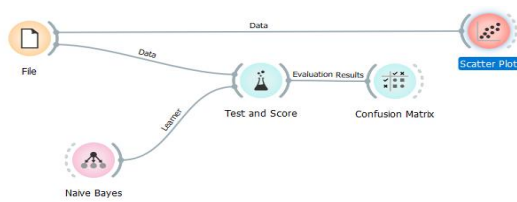
Studi ini dibagi menjadi empat tahap: pengumpulan data, pengujian data, membuat prediksi, menilai kinerja, dan membandingkan hasil dari berbagai metode. Pengumpulan dataset adalah tahap awal. Pengembangan kontribusi dan tujuan penelitian datang pertama. Metode kedua adalah pengujian data, yang melibatkan pengujian data saat ini menggunakan program-program yang membantu untuk mengumpulkan data sebagai sumber untuk klasifikasi data. Memprediksi hasil pengujian data datang pada tahap ketiga. Langkah keempat adalah menilai seberapa baik setiap teknik yang menghasilkan prediksi—KNN, Decision Tree, Random Forest Tree, Naïve Bayes, Support Vector Machine (SVM), dan AdaBoost—berperforma. Langkah keenam melibatkan pemeriksaan dan interpretasi hasil dari perbandingan teknik.

HASIL DAN PEMBAHASAN

Prosedur pengumpulan data untuk klasifikasi adalah sebagai berikut. Kumpulan data kategorisasi pendapatan, yang dapat ditemukan di situs web Kaggle, digunakan untuk menampilkan informasi tentang nilai yang diharapkan dari sebuah kelompok atribut. Kita dapat melihat tabel data dataset dalam Gambar 2. Dengan menghubungkan Preprocessing dengan KNN, Sampel Data, dan Prediksi, kita dapat menggunakan pendekatan KNN untuk mendapatkan prediksi.

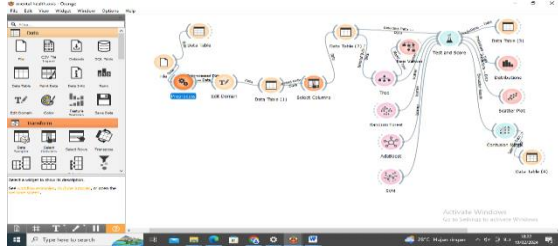


Gambar 2. Alur Kerja Algoritma KNN



Gambar 3. Algoritma Naive Bayes

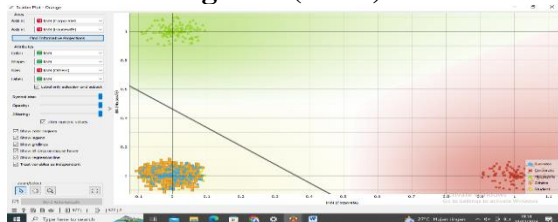
Untuk menghitung tingkat keberhasilan dan hasil Uji dan Skor, Sampel Data dari Gambar 3 dihubungkan ke Tabel Data, Naive Bayes, dan skor uji



Gambar 4. Alur Kerja Algoritma Pohon Keputusan dan SVM

Untuk melihat hasilnya, Sampel Data terhubung ke Kolom pada Gambar 4 untuk secara manual memilih domain data. Tabel Data juga ditambahkan, dan terhubung ke Pohon Viewer, Tabel Data, Hutan Acak, AdaBoost, Distribusi Plot Pencar, dan Matriks Kebingungan.

K-Nearest Neighbor (KNN)

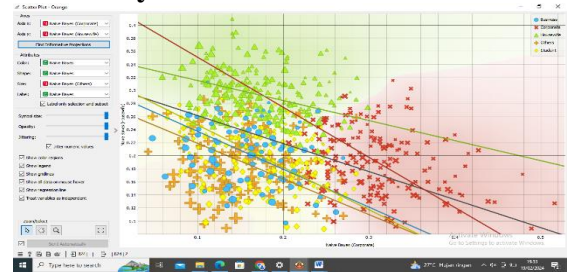


Gambar 5. K-Nearest Neighbor (KNN)

Gambar 5 menampilkan AUC sebesar 79,6%, CA sebesar 52,8%, F1 sebesar 52,2%, dan recall sebesar 52,8% untuk model KNN. Ini menunjukkan bahwa model prediksi KNN menghasilkan prediksi yang akurat, seperti yang dibuktikan oleh AUC sebesar 79,6%, atau hasil di atas 50%. Ini menunjukkan bahwa analisis pendapatan model kNN membuat prediksi yang akurat. Menurut plot pencar di atas, terdapat korelasi antara dua variabel ini pada kesehatan mental: orang yang berusia antara 25 dan 30 tahun

membentuk kategori pekerjaan dan jenis kelamin.

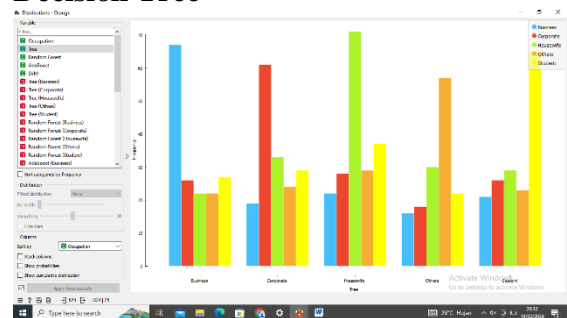
Naive Bayes



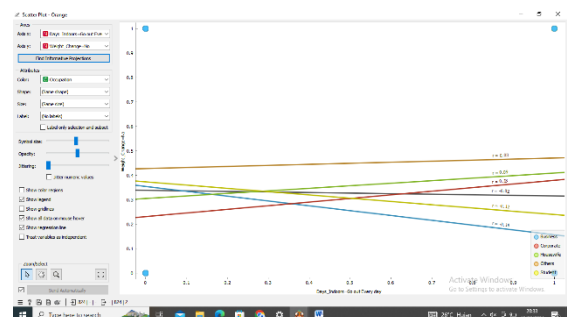
Gambar 6. Naive Bayes

Dari Gambar 6 Naive Bayes, AUC sebesar 0,620, CA sebesar 0,305, F1 sebesar 0,301, Presisi sebesar 0,305, dan Recall sebesar 0,305 ditampilkan untuk model Naive Bayes. Ini menunjukkan bahwa terdapat lebih dari 50% akurasi dalam memprediksi atau mengkategorikan usia berdasarkan kesehatan mental. Hubungan antara dua variabel ini—kesehatan mental dan jenis kelamin—terlihat pada plot pencar di atas, di mana peserta antara usia 25 dan 30 tahun merupakan mayoritas sampel.

Decision Tree



Gambar 7. Distribusi Pohon Keputusan



Gambar 8. Plot Pencar Pohon Keputusan

Pohon merupakan model algoritma terbaik dari empat yang digunakan: Hutan Acak, SVM, AdaBoost, dan Pohon. Hal ini

karena Pohon memiliki AUC terbesar, yaitu 0,736, atau 73,6%. Hasil untuk model algoritma Hutan Acak adalah 0,729, atau 72,9%, untuk model algoritma pohon adalah 0,709, atau 70,9%, dan 0,562, atau 56,2%, untuk model metode AdaBoost, yang merupakan yang terendah.

Hasil Perbandingan Model

Tabel 1 menyajikan perbandingan hasil uji berdasarkan pengujian berbagai model algoritma.

Tabel 1. Hasil Perbandingan Model

Model	AUC	CA	F1	Prec	Recall
kNN	0,796	0,528	0,522	0,551	0,528
Naïve Bayes	0,620	0,305	0,301	0,305	0,305
Tree	0,736	0,390	0,389	0,390	0,390
Random Forest	0,729	0,387	0,383	0,393	0,387
SVM	0,709	0,390	0,386	0,397	0,390
AdaBoost	0,562	0,312	0,310	0,326	0,312

Model kNN memiliki AUC sebesar 0,796, CA sebesar 0,528, F1 sebesar 0,522, Presisi sebesar 0,551, dan Recall sebesar 0,528, seperti yang dapat ditunjukkan dalam Tabel 1. Ini menunjukkan bahwa akurasi prediksi atau klasifikasi lebih tinggi dari 79%.

SIMPULAN

Dari hasil analisis berdasarkan hasil dari berbagai model mulai dari kNN, Gullible Bayes, Pohon Keputusan, Hutan Acak, Mesin Vektor Dukungan (SVM), dan AdaBoost, dapat disimpulkan bahwa model perhitungan terbaik untuk mengukur pendapatan adalah menggunakan model kNN. Hal ini dapat dilihat dari hasil AUC, CA, F1, Presisi, dan Recall yang memiliki nilai di atas 0,5 atau di atas 50%, yang menunjukkan bahwa hasil prediksi dari model kNN baik. Juga terlihat dalam hasil AUC dari model kNN dengan angka 0,796 dan hasil presisi 0,551, yang menunjukkan bahwa prediksi dari model kNN baik dan dapat dikatakan akurat. Hasil ini dapat digunakan untuk memprediksi tingkat kecemasan dan kesehatan mental yang dirasakan oleh peserta berusia 25-30 tahun, jika dianalisis berdasarkan faktor Pekerjaan dan Jenis Kelamin.

DAFTAR PUSTAKA

- [1] World Mental Health Day: An Opportunity to Kick-Start a Massive Scale-Up in Investment in Mental Health. Available online: <https://www.who.int/news/item/27-08-2020-world-mental-health-day-an-opportunity-to-kick-start-a-massive-scale-up-in-investment-in-mental-health> (accessed on 20 February 2022).
- [2] World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates. Available online: <https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf> (accessed on 20 February 2022)
- [3] Pieper, J. dan Uden, M.V. 2006. Religion in Coping and Mental Health Care. New York: Yord University Press, Inc
- [4] Dachi, J. (2023). Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit.
- [5] Giri, GA (2018). Klasifikasi Musik Berdasarkan Genre dengan Metode K-Nearest Neighbor. Jurnal Ilmu Komputer VOL. XI No.2
- [6] Indriyawati, H., & Khoirudin. (2019). Penerapan Metode Regresi Linier Dalam Koherensi Pengolahan Data Bahan Baku Tiandra Store Guna Meningkatkan Mutu Produksi. Sintak Prosiding, <https://www.unisbank.ac.id/ojs/index.php/sintak/article/view/7603>
- [7] Sau, A., Bhakta, I. (2017)"Predicting anxiety and depression in elderly patients using machine learning technology. "Healthcare Technology Letters 4 (6): 238-43.
- [8] Tyshchenko, Y. (2018)"Depression and anxiety detection from blog posts data."Nature Precis. Sci., Inst.

- Comput. Sci., Univ. Tartu, Tartu, Estonia.
- [9] Sau, A., Bhakta, I. (2018) "Screening of anxiety and depression among the seafarers using machine learning technology." *Informatics in Medicine Unlocked* :100149.
- [10] Saha, B., Nguyen, T., Phung, D., Venkatesh, S. (2016) "A framework for classifying online mental health-related communities with an interest in depression." *IEEE journal of biomedical and health informatics* 20 (4): 1008-1015.
- [11] Reece, A. G., Reagan, A. J., Lix, K. L. M., Dodds, P. S., Danforth, C. M., Langer, E. J. (2016) "Forecasting the Onset and Course of Mental Illness with Twitter Data." *Scientific reports* 7 (1): 13006.
- [12] Braithwaite, S. R., Giraud-Carrier, C., West, J., Barnes, M. D., Hanson, C.L. (2016) "Validating machine learning algorithms for Twitter data against established measures of suicidality." *JMIR mental health* 3 (2): e21.
- [13] Du, J., Zhang, Y., Luo, J., Jia, Y., Wei, Q., Tao, C., Xu, H. (2018) "Extracting psychiatric stressors for suicide from social media using deep learning." *BMC medical informatics and decision making* 18 (2): 43
- [14] Al Hanai, T., Ghassemi, M. M., Glass, J.R. (2018) "Detecting Depression with Audio/Text Sequence Modeling of Interviews." In *Interspeech*: 1716-1720.
- [15] Young, C., Harati, S., Ball, T., Williams, L. (2019) "Using Machine Learning to Characterize Circuit-Based Subtypes in Mood and Anxiety Disorders." *Biological Psychiatry* 85 (10): S310.
- [16] Liu, X. Y., Wu, J., Zhou, Z. H. (2009) "Exploratory undersampling for class-imbalance learning." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39 (2): 539-50.
- [17] Dietterich, T. G. (2000) "Ensemble methods in machine learning." In *International workshop on multiple classifier systems* Springer, Berlin, Heidelberg: 1-15.