

Pemodelan Pembelajaran Mesin untuk Prediksi Kesehatan Mental di Tempat Kerja

¹Firmansyah, ² Agus Yulianto

¹Universitas Binasarana Informatika, Indonesia, ²Universitas Nusa Mandiri

¹firmansyah.fmh@bsi.ac.id, ² agus.aag@nusamandiri.ac.id

ABSTRAK

Kesehatan mental saat ini menjadi fokus di bidang kesehatan dimana penelitian untuk mempelajari dampak lingkungan kerja terhadap kesehatan mental karyawan, serta strategi untuk meningkatkan kesejahteraan mental di tempat kerja. Meningkatnya kasus gangguan kesehatan mental dapat dicegah dengan tindakan preventif dengan memprediksi kesehatan mental pekerja dengan menggunakan beberapa model pembelajaran mesin seperti decision tree, logistic regression, naive bayes, gradient boost dan banyak metode lainnya. Dari model yang sudah diuji, dapat terlihat akurasi model Gradient Boost paling akurat 81.17% dibanding model yang lain. Model Gradient Boost dapat digunakan untuk memprediksi kesehatan mental di tempat kerja.

Kata Kunci: Kesehatan Mental, Prediksi Kesehatan Mental, Mental Health, Machine Learning, Gradient Boost, Adaboost, Decision Tree, Logistic Regression

PENDAHULUAN

Di zaman modern ini, kesadaran tentang pentingnya kesehatan mental semakin meningkat. Banyaknya tekanan dari berbagai aspek kehidupan seperti pekerjaan, hubungan, dan tuntutan sosial, serta dampak teknologi dan media sosial, semakin menekankan pentingnya kesadaran akan kesehatan mental.

Kesehatan mental dapat didefinisikan sebagai tidak adanya penyakit mental atau dapat juga didefinisikan suatu keadaan yang mencakup faktor biologis, psikologis atau sosial yang berkontribusi pada keadaan mental individu dan kemampuan untuk berguna pada lingkungan sekitar (Kementerian Kesehatan RI, 2018). Menurut sumber dari Riset Kesehatan Dasar (Riskesdas) tahun 2018 oleh Kementerian Kesehatan Republik Indonesia, terdapat sekitar 9,8% atau sekitar 26 juta dari 267 juta jiwa di Indonesia hidup dengan gangguan mental emosional atau kondisi gangguan kesehatan jiwa (Manwell LA, et al, 2015).

Untuk mencegah terjadinya gangguan kesehatan mental, tentunya perlu pencegahan dan memprediksi apakah individu memiliki potensi gangguan kesehatan mental atau tidak, dimana hal ini dapat menggunakan model pembelajaran mesin agar lebih cepat dan akurat. Model pembelajaran mesin yang digunakan merupakan algoritma yang dapat memprediksi seperti logistic regression, decision tree, random forest, gradient boost atau XGBoost.

TINJAUAN PUSTAKA

Penelitian Terkait

Sejak tahun 2013 hingga tahun 2022 telah dilakukan penelitian bahwa telah diidentifikasi terdapat 361 artikel di internet yang ditulis oleh para peneliti, dimana artikel ini menggunakan kata kunci yang berkaitan dengan AI, ML dan mental health. Dari semua artikel itu diidentifikasi terdapat artikel yang membahas ADHD sebanyak 16%, Schizofrenia sebanyak 15%, Anxiety/Depression 31%, Bipolar Disorder 9%, PTSD 15% dan Anorexia Nervosa 15% (Iyortsun

et al., 2023). Dalam rentang waktu mulai tahun 2016 sampai dengan 2020 ditemukan banyak penelitian kesehatan mental yang diambil dari konten media sosial seperti instagram, facebook, twitter, reddit dan weibo, dimana terdapat sejumlah 450 artikel pembelajaran mesin yang membahas kesehatan mental di media sosial (Kim et al., 2021).

Kesehatan Mental

Menurut organisasi kesehatan dunia WHO, kesehatan mental adalah keadaan dari mental yang sejahtera yang memungkinkan seseorang untuk mengatasi tekanan hidup, menyadari kemampuannya, belajar dengan baik dan bekerja dengan baik serta dapat berkontribusi pada komunitasnya (WHO, 2022).

Hingga tahun 2019, 1 dari 8 orang di seluruh dunia mengalami masalah kesehatan mental dan 71% dari pengidap gangguan mental tidak mendapatkan layanan kesehatan mental. Jika dilihat lebih detail, 970 juta orang di seluruh dunia hidup dengan gangguan kesehatan mental dimana 31% mengalami anxiety disorder, 28.9% mengalami depressive disorder dan sisanya seperti ADHD, bipolar disorder, autism, schizophrenia dan gangguan mental lainnya (World Health Organization, 2022).

Logistic Regression

Logistic regression atau regresi logistik adalah metode analisis statistika yang bertujuan mendeskripsikan hubungan antara variabel terikat yang memiliki dua kategori atau lebih dengan satu lebih peubah bebas yang memiliki skala kategorikal atau kontinu (Hosmer & Lemeshow, 2000). Adapun persamaan yang digunakan adalah :

$$f(y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

Dimana :

π = peluang kejadian ke-i

y_i = peubah acak ke-i yang terdiri dari 0 dan 1

1. Decision Tree

Decision tree adalah flowchart yang memiliki struktur pohon dimana masing-masing node nya memiliki atribut, setiap cabang menggambarkan keluaran dari pengujian dan setiap daun memiliki label (Han et al., 2012). Decision Tree merupakan sekian dari banyak metode di dalam data mining, dimana data mining adalah proses mengekstrak data dengan menggunakan metode dan algoritma tertentu untuk menghasilkan informasi yang lebih berguna sehingga dapat menjadi dasar dalam pengambilan keputusan (Firmansyah & Yulianto, 2021). Persamaan yang digunakan di dalam decision tree yaitu :

$$Entropy(S) = - \sum_{i=1}^n P_i * \log_2 P_i$$

Dimana :

S = himpunan kasus

n = jumlah partisi |S|

Pi = proporsi dari |Si| terhadap |S|

$$InformationGain(S, A) = Entropy(S) - \sum_{i=1}^n * Entropy(S_i)$$

Dimana :

S = himpunan kasus

A = atribut

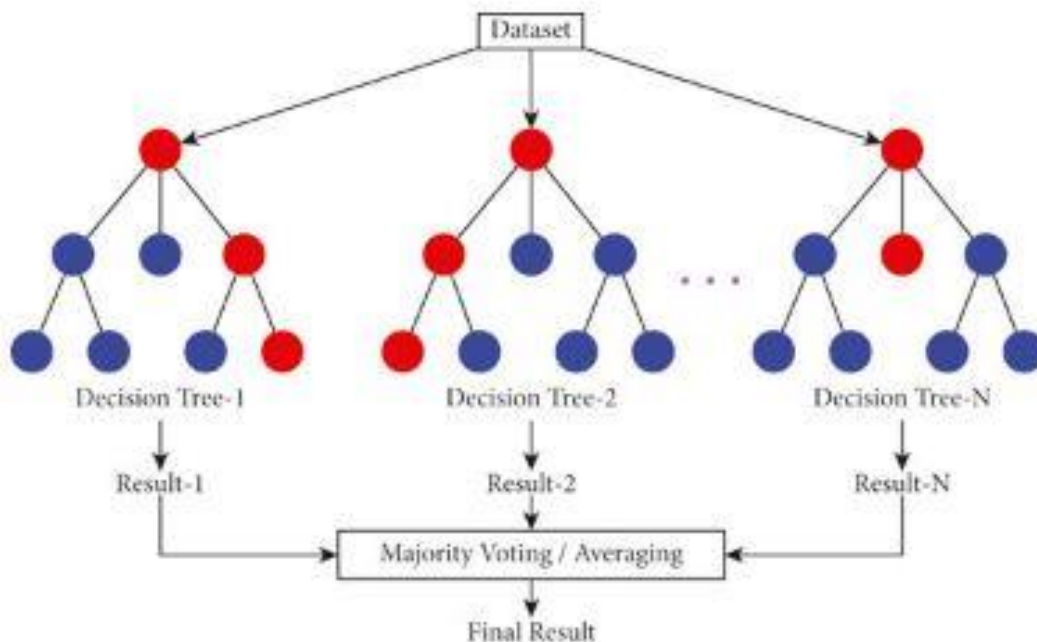
n = jumlah partisi dari atribut A

|Si| = jumlah kasus pada partisi ke-i

|S| = jumlah kasus dalam S

2. Random Forest

Random Forest adalah model yang dapat membentuk sejumlah pohon (tree) dimana setiap pohon yang dihasilkan berasal dari sejumlah data training. Setiap pohon yang dibentuk menghasilkan prediksi kelas dan prediksi kelas dengan vote terbanyak (majority vote) yang akan menjadi acuan untuk model prediksi (Breiman, 2021).



Gambar 1. Model Random Forest

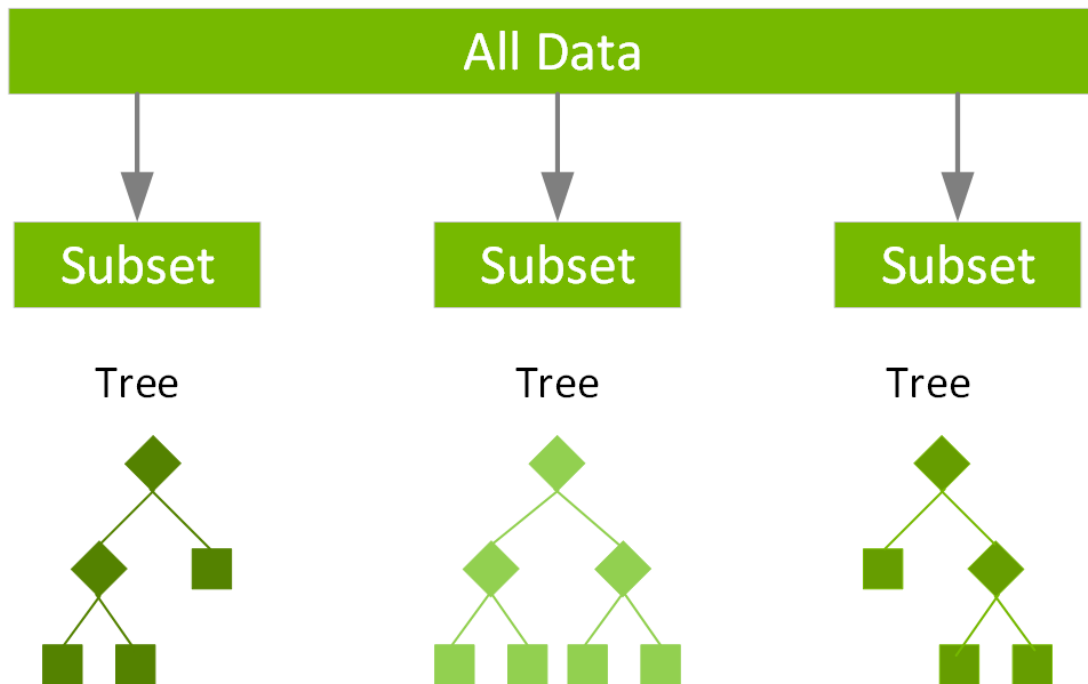
3. Gradient Boost

Gradient boost merupakan salah satu metode di dalam decision tree yang dilatih secara berurutan dimana di dalam setiap iterasi gradient boost mempelajari pohon keputusan dengan menyesuaikan gradient negative (yang dikenal dengan residual error) (Machado et al., 2019).

$$-\log L1 = - \sum_{i=1}^N y_i \log(odds) + \log(1 + e^{\log(odds)})$$

4. XGBoost

Merupakan salah satu metode boost yang berbasis decision tree untuk mempercepat kinerja dan akurasi dari model. XG Boost pertama kali dibuat oleh Tianqi Chen yang sudah tersedia dalam versi kode sumber terbuka.



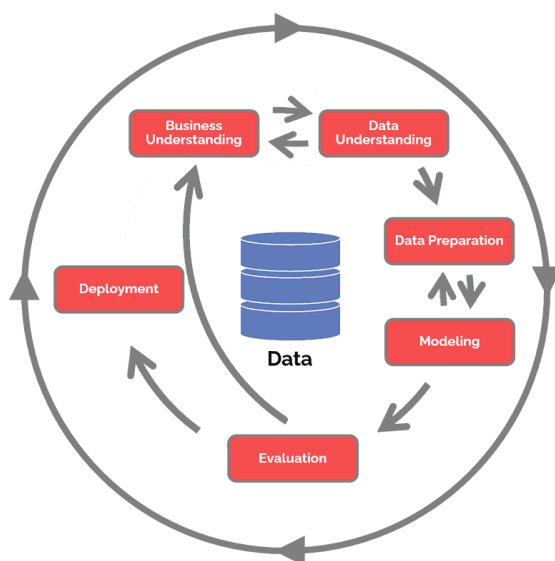
Gambar 2. Model XG Boost

METODE PENELITIAN

A. CRISP-DM (Cross-Industry Standard Process for Data Mining)

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah sebuah metode atau kerangka kerja yang digunakan untuk mengorganisasi dan mengelola proses data mining. Metode ini dibuat untuk memberikan struktur yang jelas dalam melakukan proyek data mining, mulai dari tahap perencanaan hingga evaluasi hasil. CRISP-DM terdiri dari enam tahap utama:

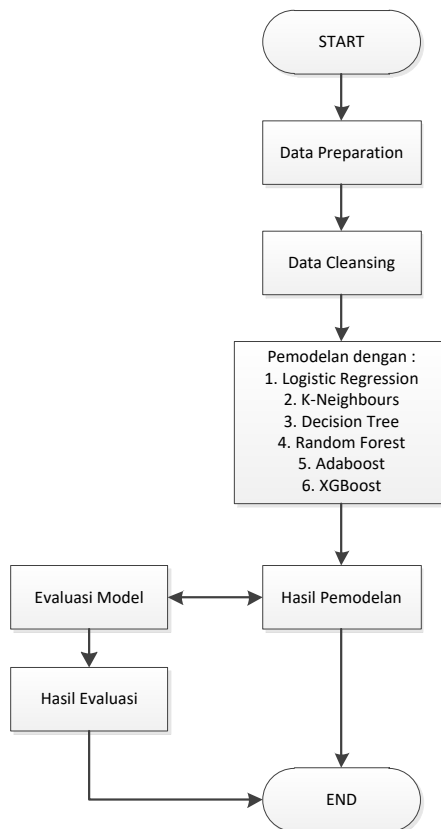
1. Understanding Business/Objective Understanding (Pemahaman Bisnis/Obyektif): Tahap ini dimulai dengan memahami tujuan bisnis atau tujuan dari proyek data mining. Fokus pada pemahaman masalah yang akan diselesaikan dan bagaimana solusi analisis data dapat memberikan nilai tambah pada bisnis.
2. Data Understanding (Pemahaman Data): Tahap ini mencakup pengumpulan data yang relevan untuk proyek, eksplorasi data, serta evaluasi kualitas data. Tujuannya adalah memahami sifat dan karakteristik data yang akan digunakan dalam analisis.
3. Data Preparation (Persiapan Data): Pada tahap ini, data yang telah dikumpulkan dan dipahami akan dipersiapkan untuk proses pemodelan. Ini mungkin melibatkan pembersihan data, transformasi variabel, penggabungan data, atau pembentukan variabel baru.
4. Modeling (Pemodelan): Tahap ini melibatkan pengembangan model analisis data menggunakan teknik seperti machine learning, statistik, atau algoritma lainnya. Tujuannya adalah menghasilkan model yang dapat memecahkan masalah atau mencapai tujuan proyek.
5. Evaluation (Evaluasi): Setelah model dibuat, tahap evaluasi digunakan untuk mengukur kinerja model dan memastikan bahwa model tersebut memenuhi tujuan bisnis yang telah ditetapkan. Evaluasi dapat dilakukan dengan menggunakan data yang terpisah atau melalui teknik validasi silang.
6. Deployment (Implementasi): Tahap terakhir adalah mengimplementasikan solusi atau model yang dihasilkan ke dalam lingkungan produksi. Ini melibatkan pengujian, pelatihan pengguna, dan integrasi solusi ke dalam infrastruktur yang ada.



Gambar 3 CRISP-DM Process

B. Flowchart Pemodelan dengan Machine Learning

Tahap awal adalah mengumpulkan data kesehatan mental yang kemudian dari data tersebut dilakukan pembersihan terhadap data agar tidak data yang inkonsisten. Setelah data dibersihkan tahap berikutnya adalah menerapkan pemodelan menggunakan beberapa model machine learning yaitu logistic regression, decision tree, random forest, gradient boost dan XGboost. :



Gambar 4 Flowcart Pemodelan

HASIL DAN PEMBAHASAN

Untuk dapat memprediksi atau mengklasifikasi maka dibutuhkan pemodelan machine learning yang memiliki tingkat akurasi tinggi dan dapat memproses data yang cukup besar. Di dalam pemodelan ini ada beberapa model yang akan diuji yaitu logistic regression, decision tree, random forest, gradient boost dan XGBoost.

Data yang diambil bersumber dari hasil survey kesehatan mental mulai dari tahun 2014 sampai dengan 2022 di perusahaan teknologi. Survey dilakukan oleh Open Sourcing Mental Illness (OSMI) sebuah organisasi nirlaba yang mendukung kepedulian dan menyediakan sumber daya kesehatan mental di amerika serikat.

Dari data yang sudah ada keseluruhan yaitu total 22 fitur, adapun masing-masing fitur itu adalah :

Table 1
Deskripsi Fitur

Fitur	Deskripsi
Age	Usia responden
Gender	Jenis kelamin
Self_employed	Pekerja lepas/freelance
Family_history	Riwayat kesehatan mental keluarga
Treatment	Apakah responden mencari pengobatan untuk kondisi kesehatan mental
Work_interfere	Kesehatan mental yang mengganggu pekerjaan
No_employees	Jumlah karyawan di kantor
Tech_company	Apakah responden bekerja di perusahaan teknologi
Benefits	Apakah perusahaan mendukung kesehatan mental
Care_options	Apakah responden tahu pilihan perawatan kesehatan yang disediakan perusahaan
Wellness_program	Apakah perusahaan pernah membahas kesehatan mental sebagai bagian dari program kesehatan mental ?
Seek_help	Apakah perusahaan menyediakan sumber daya untuk belajar kesehatan mental dan cara mengatasinya ?
Anonymity	Apakah anonimitas terlindungi jika responden jika memanfaatkan sumber daya kesehatan mental
Leave	Seberapa mudah untuk perusahaan mengijinkan untuk cuti perawatan kesehatan mental
Mentalhealthconsequence	Apakah mendiskusikan masalah kesehatan mental akan menimbulkan kesan negative ?
Physhealthconsequence	Apakah mendiskusikan masalah kesehatan fisik akan menimbulkan kesan negative ?
Coworkers	Apakah responden bersedia mendiskusikan masalah kesehatan mental dengan rekan kerja ?
Physhealthinterview	Apakah responden akan menyampaikan masalah kesehatan fisik kepada calon pemberi kerja dalam wawancara ?
Mentalvsphysical	Apakah responden merasa perusahaan menganggap kesehatan mental dan kesehatan fisik sama pentingnya?
Obs_consequence	Pernahkah Anda mendengar atau mengamati dampak negatif terhadap rekan kerja dengan kondisi kesehatan mental di tempat kerja Anda?

Dari 20 fitur, 1 kelas adalah fitur HeartDiseaseorAttack dimana fitur sudah diklasifikasikan menjadi 0 dan 1, 0 adalah tidak dan 1 adalah ya. Untuk penjelasan dari klasifikasi fitur adalah dijelaskan tabel di bawah ini :

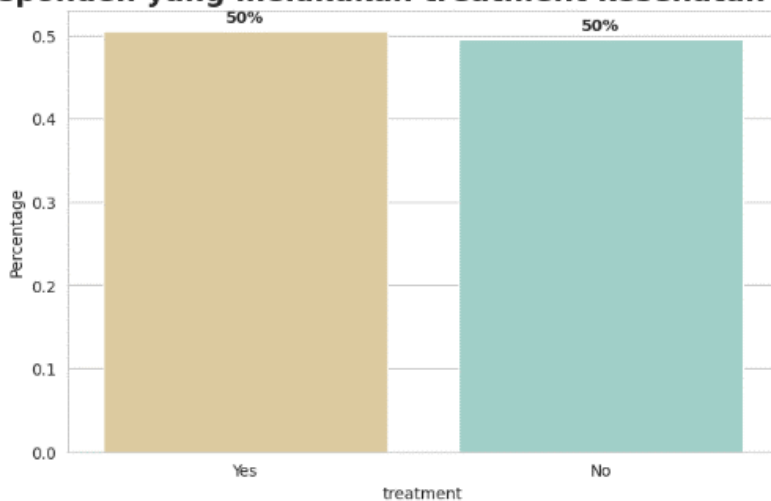
Table 2
Klasifikasi Fitur

Fitur	Klasifikasi
Age	-
Gender	Male/Female
Self_employed	Yes/No
Family_history	Yes/No
Treatment	Yes/No
Work_interfere	Never/ Often/ Rarely/ Sometimes
No_employees	-
Tech_company	Yes/No
Benefits	Yes/No/Don't know
Care_options	Yes/No/Not sure
Wellness_program	Yes/No/Don't know
Seek_help	Yes/No/Don't know
Anonymity	Yes/No/Don't know
Leave	Very easy/ Somewhat easy/ Somewhat difficult/ Very difficult/ Don't know
Mentalhealthconsequence	Yes/No/Maybe
Physhealthconsequence	Yes/No/Maybe
Coworkers	Yes/No/Some of them
Physhealthinterview	Yes/No/Maybe
Mentalvsphysical	Yes/No/Don't know
Obs_consequence	Yes/No

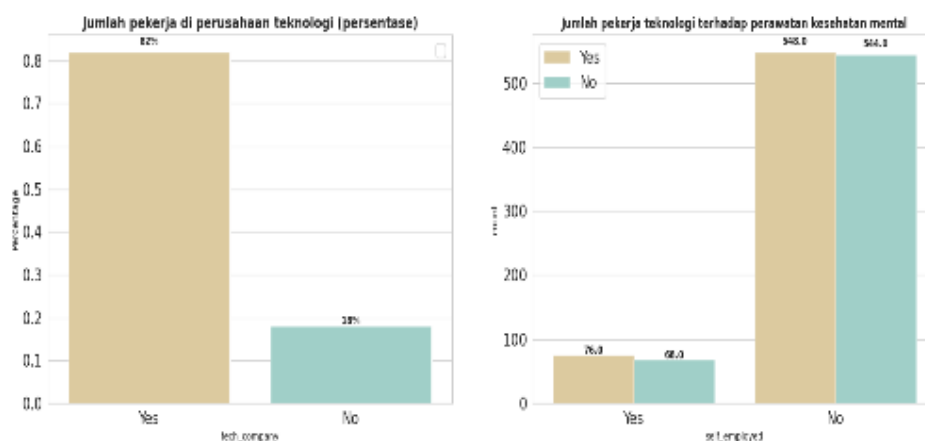
Dari sumber data yang ada, terdapat beberapa baris data yang terisi dengan nilai NA dan Null sehingga data perlu dibersihkan.

Dari data yang sudah dilakukan preprocessing didapatkan bahwa pekerja yang melakukan perawatan kesehatan mental sebesar 50% dan yang tidak melakukan sebanyak 50%.

Responden yang melakukan treatment kesehatan mental

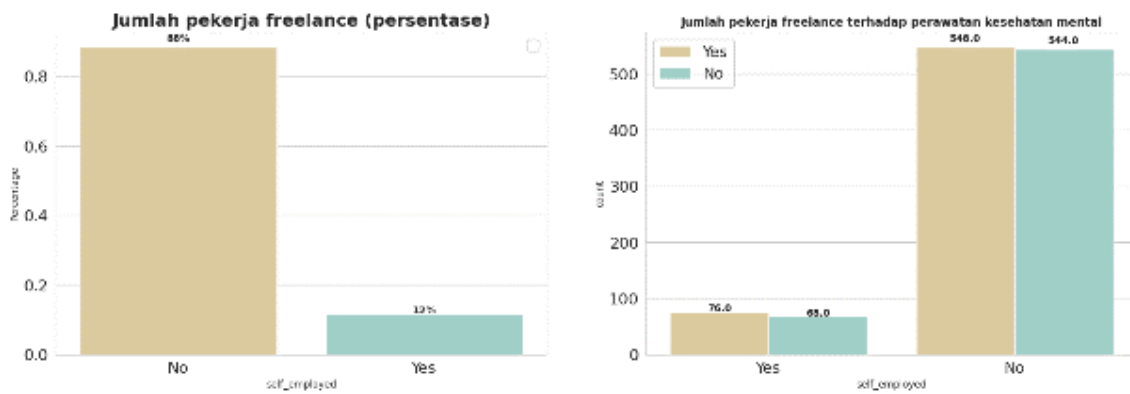


Gambar 4. Responden yang melakukan perawatan kesehatan mental dan tidak



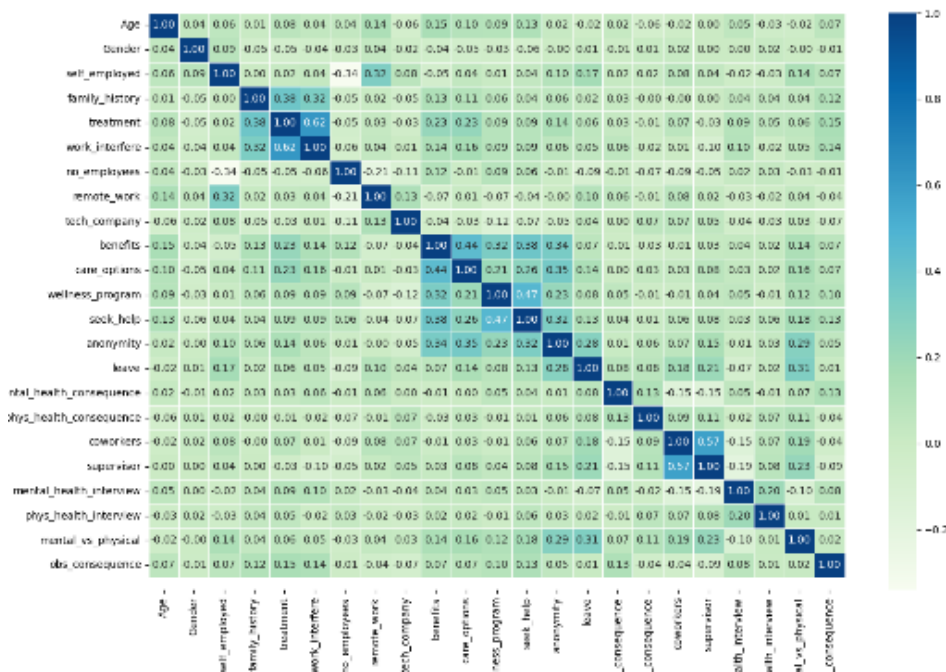
Gambar 5. Pekerja bidang teknologi yang melakukan perawatan kesehatan mental.

Dari grafik di atas juga ditemukan fakta bahwa sebanyak 82% responden merupakan pekerja di perusahaan teknologi dan dari 82% tersebut, sebanyak 511 orang diantaranya menjalankan perawatan kesehatan mental. Artinya disini dapat dilihat bahwa pekerja di bidang teknologi sebagian besar mengalami gangguan kesehatan mental. Grafik di bawah ini juga memperlihatkan bahwa pekerja bebas (freelance) lebih sedikit mengalami gangguan kesehatan mental.



Gambar 6. Pekerja lepas yang melakukan perawatan kesehatan mental

Jika kita menggunakan correlation matrix dengan grafik heatmap, dapat dilihat ada kerkaitan erat antara fitur seek_help dan fitur wellness_program, fitur supervisor dan fitur coworker, fitur care_options dan fitur benefit.



Gambar 7. Correlation matrix

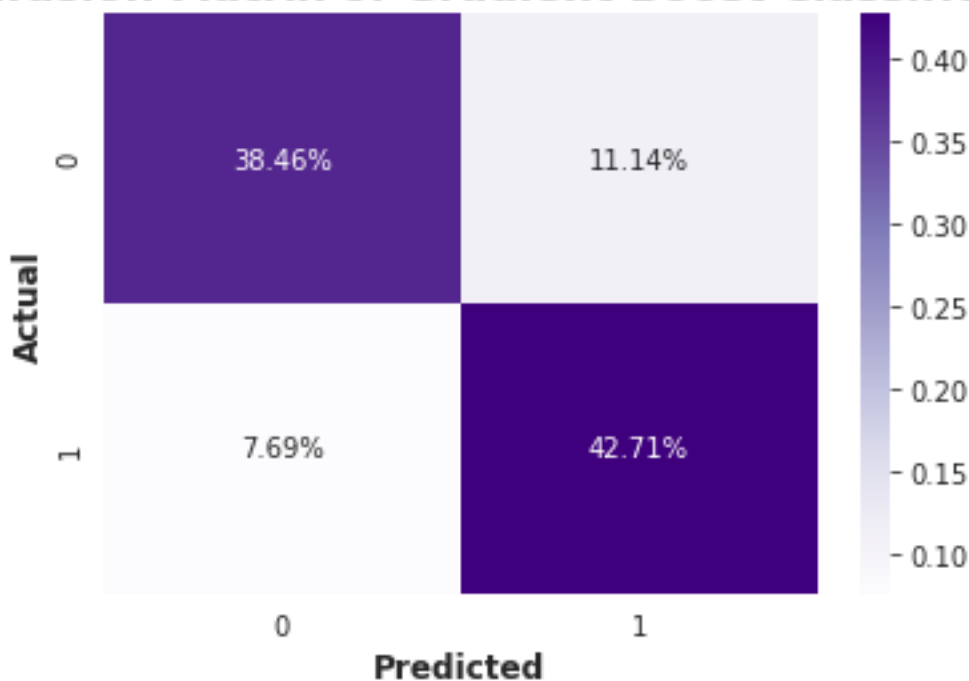
Dari hasil evaluasi model dapat dilihat model Gradient Boost classifier menghasilkan akurasi yang paling tinggi sebesar 0.8116 atau 81.16% disusul kemudian adalah model Logistic Regression.

Table 3.
Akurasi Model Algoritma

Model	Akurasi
Logistic Regression	0.8037
K-Neighbours	0.6631
Decision Tree	0.7214
Random Forest	0.7771
Gradient Boost	0.8116
Ada Boost	0.8010
XGBoost	0.7931

Hasil evaluasi menggunakan confusion matrix untuk model Gradient Boost dapat dilihat nilai True Positive sebesar 42.72 % dan True Negative 36.46% sehingga menghasilkan nilai akurasi mencapai 81.17%, presisi sebesar 79% dan recall sebesar 84%.

Confusion Matrix of Gradient Boost Classifier



Gambar 8. Confusion matrix

KESIMPULAN

Dari hasil evaluasi diatas, Gradient Boost terbukti dapat memprediksi kesehatan mental sebesar 81.17% dibandingkan model yang lain. Model yang lain seperti Adaboost dan Logistic regression juga memiliki tingkat akurasi yang tidak terpaut jauh. Dengan pemodelan ini, maka akan membantu berbagai pihak seperti perusahaan, instansi kesehatan hingga pemerintah agar lebih mudah mendeteksi kesehatan mental lebih dini.

REFERENSI

- Breiman, L. (2021). Random Forest. *Machine Learning*, 45, 5–32. https://doi.org/10.1007/978-3-030-62008-0_35
- Firmansyah, & Yulianto, A. (2021). Market Basket Analysis for Books Sales Promotion using FP Growth Algorithm, Case Study : Gramedia Matraman Jakarta. *Journal of Informatics and Telecommunication Engineering*, 4(January), 383–392.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (3rd ed.). Elsevier Inc.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). John Wiley and Sons Inc.
- Iyortsuun, N. K., Kim, S. H., Jhon, M., Yang, H. J., & Pant, S. (2023). A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis. *Healthcare (Switzerland)*, 11(3), 1–27. <https://doi.org/10.3390/healthcare11030285>
- Kim, J., Lee, D., & Park, E. (2021). Machine learning for mental health in social media: Bibliometric study. *Journal of Medical Internet Research*, 23(3), 1–17. <https://doi.org/10.2196/24870>
- Machado, M. R., Karray, S., & De Sousa, I. T. (2019). LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. *14th International Conference on Computer Science and Education, ICCSE 2019, Nips*, 1111–1116. <https://doi.org/10.1109/ICCSE.2019.8845529>
- WHO. (2022). *Mental Health*. <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>
- World Health Organization. (2022). *World mental health report*.
- Kementrian Kesehatan RI. Laporan Riset Kesehatan Dasar 2018. Jakarta; 2018.
- Manwell LA, Barbic SP, Roberts K, Durisko Z, Lee C, Ware E, et al. What is Mental Health? Evidence Towards a New Definition from a Mixed Methods Multidisciplinary International Survey. *BMJ Open*. 2015;5(6):1–11.