

Predicting University Students Mental Health Conditions with Classifications Model

Rapid Miner Models

Rudy Kurniawan Efendy

*Computer Science Department, School of
Computer Science*

*Bina Nusantara University
Tangerang, Indonesia 15143
rudy.efendy@binus.ac.id*

Juan Felix

*Computer Science Department, School of
Computer Science*

*Bina Nusantara University
Tangerang, Indonesia 15143
juan.felix001@binus.ac.id*

Priscilla Yo

*Computer Science Department, School
of Computer Science*

*Bina Nusantara University
Tangerang, Indonesia 15143
priscilla.yo@binus.ac.id*

Andrew William Corputty

*Computer Science Department, School of
Computer Science*

*Bina Nusantara University
Tangerang, Indonesia 15413
andrew.corputty@binus.ac.id*

Abstract— Data mining adalah proses mencari pola dari kumpulan data besar yang nantinya diambil ilmu pengetahuannya dan dibuat model yang dapat dimengerti dan terstruktur. Pada penelitian ini, kami melakukan prediksi kesehatan mental mahasiswa dengan classification model, mengingat semakin meningkatnya jumlah mahasiswa yang mempunyai gangguan mental. Kami menggunakan berbagai algoritma seperti Decision Tree, KNN, dan Naive Bayes untuk membandingkan hasil yang dihasilkan masing-masing algoritma tersebut. Kriteria yang digunakan untuk mengukur kualitas algoritma tersebut bisa dilihat dari akurasi, kecepatan, skalabilitas, dan lain-lain. Untuk proses data mining ini akan digunakan software Rapid Miner.

Keywords— *mental health, data mining, classification model, KNN, decision tree, naive bayes, rapid miner.*

I. INTRODUCTIONS

Akhir-akhir ini kita sering sekali mendengar istilah mental health. Dimana kita diminta untuk melakukan pengecekan terhadap mental health kita. Namun kita belum mengetahui secara pasti apa itu mental health dan mengapa itu sangat penting untuk diri kita. Menurut UU nomor 18 tahun 2014 mental health adalah sebuah kondisi dimana seseorang dapat berkembang secara fisik, mental dan spiritual sehingga orang tersebut dapat mengetahui batas dan kemampuannya sendiri dan dapat mengatasi tekanan dimana apabila kita tidak menjaga mental health kita, kita akan mengalami sedikit masalah dalam melakukan aktivitas kita dimana salah satu tanda bahwa kita memiliki masalah dengan mental health kita

menurut journal of health and therapy beberapa tanda adalah ketika kita mengalami depresi, sering kali merasa cemas baik itu cemas yang kronis maupun akut[1]. dimana ketika kita mengalami gejala gejala tersebut maka kita harus bahwa kita memiliki masalah dengan mental health kita dan mencari pertolongan seperti psikologis dan tidak mengabaikan masalah mental health tersebut. karena kita sering kali mengabaikan kesehatan mental health kita karena kita merasa bahwa kita baik baik saja padahal kita sedang tidak baik baik saja. dimana salah satu alasan mengapa orang tidak mencari pertolongan adalah rasa malu yang mereka miliki apabila menceritakan masalah tersebut kepada orang lain yang menyebabkan orang tersebut hanya menyimpan perasaan itu sendiri dan tidak mencari pertolongan padahal itu adalah tindakan yang sangat berbahaya dimana apabila tidak ditangani dengan baik maka mental health juga dapat mengakibatkan dampak dampak negatif kepada hidup kita seperti ketika kita mengalami depresi kita akan menjadi tidak termotivasi untuk melakukan apa saja sehingga membuat kita tidak produktif dan bila tidak ditangani kita akan semakin parah yang dapat mengakibatkan kematian yang disebabkan oleh bunuh diri, selain bunuh diri kita juga dapat terjerumus ke dalam berbagai jenis obat obat terlarang yang membuat kita berujung kematian karena overdosis.

Oleh sebab itu kami ingin mencegah hal hal tersebut terjadi kepada kita sehingga kita ingin mencari tahu apakah seseorang itu mental health yang baik atau tidak dengan menggunakan dataset yang kami ambil dari kaggle dimana dataset kami mengambil data sejumlah mahasiswa yang memiliki beberapa gejala dan kami ingin mencari tau dari dataset tersebut apakah mereka memiliki mental health yang sehat atau tidak[2]. dimana kami menggunakan rapid miner dengan metode knn..

II. LITERATURE REVIEW

A. University Students Mental Health

Mental health atau kesehatan mental menurut WHO bukan hanya sekedar tidak memiliki gangguan mental, melainkan suatu keadaan atau kondisi yang dimana seseorang dapat menggali potensinya sendiri, dapat mengelola stress dalam kehidupan sehari-hari, dapat bekerja secara produktif, serta mampu memberikan kontribusi pada masyarakat[3]. Ketika membicarakan kesehatan mental mahasiswa, seringkali konsep kesehatan mental yang disampaikan WHO ini tidak terealisasi atau dialami oleh kalangan mahasiswa. Seiring berjalannya waktu, semakin banyak mahasiswa yang memiliki gangguan kesehatan mental. Hal tersebut sangat terlihat saat pandemi COVID-19 dimana terjadi peningkatan gangguan mental pada mahasiswa sebanyak 2 kali lebih tinggi pada tahun 2020 dibandingkan dengan tahun 2019. Penelitian ini dilakukan oleh Student Experience in the Research University (SERU) melalui survey pada 30.725 mahasiswa sarjana dan 15.346 mahasiswa pascasarjana dan profesional[4].

B. Data Mining

Data mining adalah proses mencari pola dari kumpulan data besar yang nantinya diambil ilmu pengetahuannya dan dibuat model yang dapat dimengerti dan terstruktur. Model tersebut dibuat dengan algoritma yang nantinya akan mempelajari training dataset untuk dibuat model dan dipakai pada data lain untuk prediksi. Hal inilah yang disebut classification dan clustering, atau sering disebut juga supervised dan unsupervised learning. Terdapat dua langkah pada klasifikasi yaitu pada langkah pertama algoritma klasifikasi menggunakan data pelatihan untuk membangun pengklasifikasi, dan kemudian pada langkah kedua pengklasifikasi ini digunakan untuk memprediksi label kelas dari contoh data yang tidak berlabel.

C. Processing Data Before Classification

Berikut adalah tahap-tahap preprocessing yang harus dilakukan untuk mendapatkan akurasi, efisiensi, dan stabilitas yang baik saat proses klasifikasi:

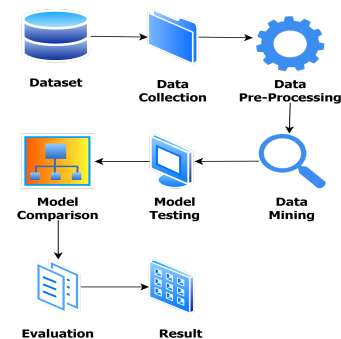
- Data Cleaning**
Proses pembersihan data dengan menghilangkan atau menghapus noise serta mengatasi nilai yang hilang.
- Relevance Analysis**
Suatu dataset mungkin saja mempunyai atribut yang berulang (redundant), sehingga bisa digunakan analisis korelasi untuk mencari tahu apakah dua atau lebih atribut saling berhubungan.
- Data Transformation**
Terkadang dataset perlu ditransformasi dengan normalisasi, terutama ketika menggunakan algoritma seperti back propagation of neural network atau K-NN. Normalisasi dilakukan dengan menyesuaikan semua nilai dari atribut yang dipertimbangkan sehingga mereka berada dalam rentang tertentu.

D. Criteria of Comparing Classification

- Accuracy**
Kemampuan classifier untuk dengan benar memprediksi label kelas dari data yang belum pernah dilihat.
- Speed**
Biaya komputasi yang terlibat dalam membangun model dan kecepatannya.
- Robustness**
Kemampuan classifier untuk membuat prediksi yang benar, bahkan jika data terdapat banyak noise dan memiliki nilai yang hilang.
- Scalability**
Kemampuan untuk membuat model klasifikasi secara efisien dari data yang besar.
- Interpretability**
Kemudahan atau sejauh mana pemahaman yang diberikan oleh classifier dapat dimengerti.

III. METHODOLOGY

A. Proposed Methodology



gambar 1. Alur kerja penelitian

Penelitian ini bertujuan untuk memprediksi apakah mahasiswa tersebut memerlukan perawatan dari spesialis terhadap masalah kesehatan mental berdasarkan dataset yang telah dikumpulkan. Penelitian ini menggunakan dataset yang akan diproses terlebih dahulu untuk menghilangkan dataset yang tidak bisa digunakan dan dataset yang kosong sehingga dataset bersih dan bisa dilanjutkan ke tahap berikutnya. Dataset yang telah bersih akan diterapkan ke tiga algoritma data mining seperti K-NN, Naive Bayes dan Decision Tree. Hasil perhitungan dari tiga algoritma tersebut akan dibandingkan sehingga terpilihnya algoritma dengan performa yang terbaik. Penelitian ini diakhiri dengan hasil dan visualisasi dari algoritma-algoritma yang telah diterapkan terhadap dataset.

B. Dataset

Dataset merupakan komponen terpenting dalam memprediksi apakah diperlukannya tenaga spesialis pada mahasiswa terkait masalah kesehatan mental. Dataset yang

digunakan pada penelitian ini adalah dataset yang diambil dari sumber online bernama kaggle. Dataset yang digunakan merupakan dataset dari mahasiswa yang diberikan survey berkaitan dengan kesehatan mental. Dataset ini terdapat 11 atribut dan dataset tersebut dikumpulkan pada tahun 2020. Berikut ini adalah atribut-atribut yang digunakan pada dataset beserta deskripsi dari setiap atribut adalah

Atribut	Deskripsi
Timestamp	Waktu yang diambil pada saat mahasiswa yang mengisi survey
Choose your gender	Jenis kelamin dari mahasiswa
Age	Umur dari mahasiswa
What is your course?	Jurusan kuliah dari mahasiswa
Your current year of Study	Tahun dari mahasiswa yang belajar di jurusan yang dipilih
What is your CGPA?	Nilai dari mahasiswa
Marital status	Status pernikahan dari mahasiswa
Do you have Depression?	Apakah mahasiswa mengalami depresi?
Do you have Anxiety?	Apakah mahasiswa mengalami kegelisahan?
Do you have Panic attack?	Apakah mahasiswa mengalami serangan panik?
Did you seek any specialist?	Apakah mahasiswa mencari spesialis?

gambar 2. Tabel atribut dataset beserta deskripsi

C. Data Mining

Penelitian ini menggunakan teknik data mining untuk memprediksi apakah seorang mahasiswa akan memerlukan tenaga spesialis terkait masalah kesehatan mental. Terdapat tiga algoritma data mining yang digunakan pada penelitian ini yaitu k-NN, Naive Bayes dan Decision Tree. Tiga algoritma ini akan digunakan untuk dibandingkan dari segi performa, akurasi, dan error.

a). K-NN

K-Nearest Neighbor atau yang disingkat sebagai K-NN adalah sebuah algoritma yang mengklasifikasikan sekelompok data berdasarkan pembelajaran data yang sudah diklasifikasikan

sebelumnya. K-NN mengelompokkan data berdasarkan kedekatan atau kemiripan data yang baru dengan data yang terdekat.

b). Naive Bayes

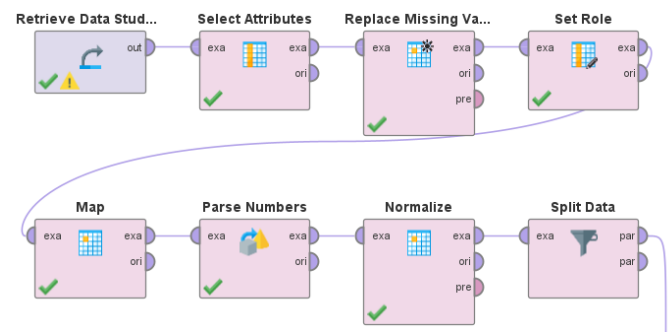
Naive Bayes adalah sebuah algoritma yang berakar pada teorema Bayes yang dimana ciri utama dari teorema ini adalah adanya asumsi yang sangat kuat terhadap independensi dari kondisi atau kejadian dari masing-masing data.

c). Decision Tree

Decision Tree adalah sebuah algoritma yang menggunakan beberapa aturan untuk membuat keputusan sehingga hasil yang didapatkan dari algoritma ini berbentuk pohon. Decision tree dimulai dengan satu node dan akan terus memanjang sehingga setiap node yang dihasilkan merupakan pilihan dari tindakan lain.

D. Data Pre-Processing

Pada penelitian ini akan menggunakan dataset yang telah diambil dari kaggle. Akan tetapi, dataset tersebut belum bisa diterapkan ke dalam algoritma karena dataset belum dibersihkan atau dataset tersebut masih terdapat kesalahan. Oleh karena itu, dibutuhkannya tindakan untuk mengatasi kesalahan-kesalahan tersebut.



gambar 3. Data Pre-Processing

Dataset yang ingin digunakan akan di import atau ditambahkan kedalam RapidMiner agar bisa diterapkan ke tiga algoritma. Dataset akan dihilangkan terlebih dahulu yang diperkirakan tidak penting dengan menggunakan operator ‘Select Attributes’ dan dari operator ini, atribut akan di exclude atau di tidak di tiadakan dari dataset. Setelah itu, untuk menghilangkan data atau nilai yang kosong atau hilang pada dataset dengan menggunakan operator ‘Replace Missing Value’. Satu atribut pada Dataset akan dijadikan sebagai label untuk bisa dijadikan acuan prediksi dengan menggunakan operator ‘Set Role’. Menggunakan operator ‘Map’ untuk mengubah isi dari setiap atribut yang sebelumnya terdapat kata menjadi angka sehingga bisa diterapkan dalam model. setelah itu menggunakan operator ‘Parse Numbers’ untuk mengubah bentuk atribut yang tadinya polynomial menjadi number.

The diagram illustrates a data analysis workflow. It begins with two initial steps: "Retrieve Data Stude..." and "Replace Missing Val...". The workflow then branches into two main paths. The upper path consists of "Map", "Parse Numbers", "Normalize", and "Split Data". The lower path consists of "Multiply". Both paths converge into three machine learning models: "k-NN (2)", "Decision Tree (2)", and "Naive Bayes (2)". Each model has multiple output ports labeled "mod", "exa", "tes", "per", and "per".

Data yang sudah dibagi atau split akan digunakan ke tiga algoritma yaitu K-NN, Naive Bayes, dan Decision Tree. Menggunakan operator 'Multiply' agar data tersebut bisa digunakan kepada tiga algoritma secara bersamaan. Menggunakan operator 'Cross Validation' untuk algoritma dapat diterapkan.

A. Dataset Preprocessing

transformasi data atau *data transformation*[5].

[illegible]

B. Result

[illegible][illegible]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

Perlu diklasifikasikan, bahwa tabel-tabel diatas merupakan sampel daripada data yang diproses sesungguhnya guna meningkatkan kejelasan dalam penyampaian informasi. Ada sekitar 100 dataset yang diproses, dan menggunakan fungsi *Split Data* pada Rapidminer, kami membagi data tersebut menjadi sekitar 30% data *training* dan 70% data *testing*. Daripada hasil yang didapatkan, kami berhasil mendapati data sebagai berikut:

Parameter	K-NN	Naive Bayes	Decision Tree
Accuracy	94.37%	70.42%	90.14%
Absolute Error	0.099	0.289	0.128
Root Mean Squared Error	0.251	0.529	0.311
Squared Error	0.063	0.280	0.097
Squared correlation	0.000	0.012	0.000

V. CONCLUSIONS

Penelitian ini bertujuan untuk menguji beberapa algoritma *machine learning* dalam memprediksi sebuah *outcome* dari data-data yang ada. Berdasarkan hasil yang diperoleh, dengan menggunakan parameter global, yakni Accuracy (Keakuratan), algoritma K-NN merupakan algoritma yang tepat untuk digunakan pada dataset ini.

REFERENCES

- [1] F. Ulya, "Literature Review Of Factors Related To Mental Health In Adolescent: Kajian Literatur Faktor Yang Berhubungan Dengan Kesehatan Mental Pada Remaja," *J. Health Ther.*, vol. 1, no. 1, Art. no. 1, Nov. 2021.
- [2] "Student Mental health." Accessed: Dec. 16, 2023. [Online]. Available: <https://www.kaggle.com/datasets/shariful07/student-mental-health>
- [3] Hernández-Torrano, D., Ibrayeva, L., Sparks, J., Lim, N., Clementi, A., Almukhambetova, A., ... & Muratkyzy, A. (2020). *Mental health and well-being of university students: A bibliometric mapping of the literature*. *Frontiers in psychology*, 11, 1226.
- [4] Chrikov, I., Soria, K. M., Horgos, B., & Jones-White, D. (2020). *Undergraduate and graduate students' mental health during the COVID-19 pandemic*.
- [5] Singh, D., & Singh, B. (2019). *Investigating the impact of data normalization on classification performance*. *Applied Soft Computing*, 105524. doi:10.1016/j.asoc.2019.105524
- [6] Choudhary, R., & Gianey, H. K. (2017). *Comprehensive Review On Supervised Machine Learning Algorithms*. *2017 International Conference on Machine Learning and Data Science (MLDS)*. doi:10.1109/mlds.2017.11
- [7] Verma, A. (2018). *Study and Evaluation of classification algorithms in data mining*. *Int. Res. J. Eng. Technol*, 5, 1-11.
- [8] Mohanapriya, M., & Lekha, J. (2018, November). *Comparative study between decision tree and knn of data mining classification technique*. In *Journal of Physics: Conference Series* (Vol. 1142, No. 1, p. 012011). IOP Publishing.