



Implementasi Algoritma Naïve Bayes dalam Pengklasifikasian Kasus Stunting pada Posyandu Mangga Jagakarsa

Ahnaf Al Latief^{1*}, Indra Kristianto²

^{1,2} Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
Email: ¹latiefahnaf@gmail.com, ²dosen02597@unpam.ac.id

Abstrak—Stunting merupakan kondisi gangguan pertumbuhan yang dapat memberikan dampak jangka panjang terhadap perkembangan anak. Kondisi ini tidak hanya berkaitan dengan pertumbuhan fisik, tetapi juga berhubungan dengan perkembangan kognitif dan pencapaian pendidikan. Penelitian longitudinal di Indonesia menunjukkan bahwa gangguan pertumbuhan pada masa kanak-kanak berhubungan dengan kemampuan kognitif dan kemampuan numerik yang lebih rendah serta pencapaian pendidikan yang lebih rendah. Metode pengumpulan data yang digunakan berfokus pada perolehan data yang akan diolah menggunakan algoritma klasifikasi. Evaluasi performa algoritma Naive Bayes dapat dilakukan dengan menggunakan beberapa ukuran evaluasi. Accuracy digunakan untuk mengukur persentase keseluruhan prediksi yang benar terhadap seluruh data yang diuji. Precision berfokus pada proporsi prediksi yang benar dari seluruh data yang diprediksi ke dalam suatu kelas tertentu, sedangkan recall digunakan untuk menilai kemampuan model dalam mengenali dan mengklasifikasikan data yang benar (true positive). Hasil pengujian model menunjukkan tingkat keberhasilan klasifikasi yang sangat tinggi, dimana nilai accuracy dan weighted mean recall yang diperoleh mencapai 84.38%. Hal ini menunjukkan bahwa algoritma Naive Bayes memiliki performa yang sangat baik dalam mengidentifikasi status stunting pada data balita di Posyandu Mangga Jagakarsa.

Kata Kunci: klasifikasi; Naïve Bayes; Data Mining; Stunting; posyandu

Abstract— *Stunting is a growth disorder that can have long-term impacts on child development. This condition is not only related to physical growth, but is also associated with cognitive development and educational attainment. Longitudinal research in Indonesia shows that growth impairment during childhood is associated with lower cognitive and numeracy abilities as well as lower educational attainment. The data collection method used focuses on obtaining data that will be processed using a classification algorithm. The performance of the Naive Bayes algorithm can be evaluated using several evaluation measures. Accuracy is used to measure the overall percentage of correct predictions against all tested data. Precision focuses on the proportion of correct predictions among all data predicted into a particular class, while recall is used to assess the model's ability to recognize and correctly classify data (true positive). The model testing results show a very high classification success rate, where the accuracy and weighted mean recall values obtained reached 84.38%. This indicates that the Naive Bayes algorithm performs very well in identifying stunting status in toddler data at Posyandu Mangga Jagakarsa.*

Keywords: *classification; Naïve Bayes; Data Mining; stunting; integrated health post (Posyandu)*

1. PENDAHULUAN

Stunting merupakan kondisi gangguan pertumbuhan yang dapat memberikan dampak jangka panjang terhadap perkembangan anak. Kondisi ini tidak hanya berkaitan dengan pertumbuhan fisik, tetapi juga berhubungan dengan perkembangan kognitif dan pencapaian pendidikan. Penelitian longitudinal di Indonesia menunjukkan bahwa gangguan pertumbuhan pada masa kanak-kanak berhubungan dengan kemampuan kognitif dan kemampuan numerik yang lebih rendah serta pencapaian pendidikan yang lebih rendah. Dampak tersebut selanjutnya dapat berimplikasi terhadap potensi pendapatan seseorang sepanjang hidup melalui penurunan kemampuan kognitif dan tingkat Pendidikan (Lestari et al., 2024).

Indonesia masih menghadapi permasalahan stunting yang menjadi salah satu tantangan dalam bidang kesehatan dan gizi masyarakat. Upaya penurunan stunting memerlukan pendekatan yang komprehensif karena faktor yang memengaruhi kejadian stunting tidak hanya berkaitan dengan asupan gizi, tetapi juga kondisi sanitasi, akses terhadap air bersih, pendidikan, serta faktor sosial dan ekonomi. Intervensi sensitif menjadi salah satu strategi yang dapat mendukung percepatan penurunan stunting melalui peningkatan akses sanitasi, air bersih, serta berbagai program lintas sektor. Penelitian di DKI Jakarta menunjukkan bahwa peningkatan fasilitas sanitasi dan akses



terhadap air bersih merupakan bagian penting dalam intervensi sensitif untuk mengurangi risiko stunting. Selain itu, penerapan intervensi sensitif di berbagai wilayah Indonesia menunjukkan bahwa keberhasilan percepatan penurunan stunting membutuhkan keterlibatan berbagai sektor dan strategi yang disesuaikan dengan kondisi masing-masing wilayah (Irawati et al., 2023).

Posyandu memiliki peran penting dalam pemantauan pertumbuhan dan status gizi balita melalui kegiatan pengukuran antropometri secara berkala. Namun, proses penentuan status stunting masih dapat dilakukan secara manual berdasarkan hasil pengukuran antropometri. Proses tersebut berpotensi menimbulkan kesalahan dalam perhitungan maupun penginputan data sehingga diperlukan suatu sistem yang dapat membantu proses klasifikasi status stunting secara lebih cepat dan terstruktur (Adzim et al., 2023).

Data mining dapat dimanfaatkan untuk membantu proses klasifikasi status stunting pada balita dengan mempelajari pola yang terdapat pada data sebelumnya. Data antropometri seperti usia, jenis kelamin, berat badan, dan tinggi badan dapat digunakan sebagai atribut dalam proses klasifikasi. Pemanfaatan metode klasifikasi memungkinkan data yang sebelumnya diolah secara manual dapat diproses secara lebih sistematis sehingga dapat membantu mengidentifikasi status stunting atau normal pada balita secara lebih efektif (Mulyono et al., 2023).

Dalam memilih algoritma klasifikasi yang optimal, pertimbangan utama adalah efisiensi komputasi dan akurasi, terutama untuk implementasi di lingkungan yang mungkin memiliki keterbatasan sumber daya hardware. Algoritma Naive Bayes dipilih karena reputasinya sebagai metode klasifikasi probabilistik yang efisien. Algoritma ini didasarkan pada Teorema Bayes, dengan asumsi utama bahwa setiap fitur dalam data (misalnya, usia, berat badan) berkontribusi secara independen terhadap probabilitas hasil akhir (stunting atau tidak stunting), terlepas dari korelasi antar fitur lainnya. Keunggulan Naive Bayes terletak pada kecepatan pelatihannya yang luar biasa, kemudahan interpretasinya, dan kemampuannya memberikan kinerja yang sangat baik dalam kasus data medis dan klasifikasi teks, bahkan ketika asumsi independensi yang "naif" tersebut tidak sepenuhnya terpenuhi (Alpaydin, 2020). Efisiensi ini menjadikannya kandidat ideal untuk diterapkan pada sistem pendukung keputusan yang membutuhkan respons cepat, seperti di lingkungan Posyandu.

2. METODE PENELITIAN

2.1 Jenis dan Sumber Data

Penelitian ini menggunakan pendekatan kuantitatif dengan objek penelitian berupa data balita pada Posyandu Mangga Jagakarsa. Data yang digunakan merupakan data balita yang diperoleh dari arsip Posyandu.

Pengumpulan data dilakukan melalui observasi, wawancara, studi literatur, dan dokumentasi. Observasi dilakukan untuk mengetahui proses pencatatan dan pemeriksaan data balita. Wawancara dilakukan dengan kader Posyandu untuk memperoleh informasi mengenai proses penentuan status pertumbuhan dan kendala pengolahan data. Dokumentasi digunakan untuk memperoleh data balita yang tersedia pada arsip Posyandu.

2.2 Variabel Input

Variabel input yang digunakan dalam penelitian ini adalah:

1. Jenis Kelamin Balita, digunakan sebagai variabel pendukung karena standar pertumbuhan balita berbeda antara laki-laki dan perempuan.
2. Umur Balita (bulan), digunakan untuk mengetahui tahapan pertumbuhan balita secara lebih akurat.
3. Berat Badan Balita, digunakan untuk menilai kondisi pertumbuhan berdasarkan indikator berat badan menurut umur.
4. Tinggi Badan Balita, digunakan sebagai indikator utama dalam menentukan status stunting.

Sebelum diproses menggunakan algoritma Naive Bayes, data tersebut terlebih dahulu diklasifikasikan ke dalam kategori tertentu berdasarkan standar antropometri WHO.



2.3 Data Mining

Knowledge Discovery in Databases (KDD) merupakan suatu proses yang digunakan untuk menemukan informasi atau pola yang menarik dan bermanfaat dari sekumpulan data dalam jumlah besar. Proses KDD terdiri dari beberapa tahapan, yaitu seleksi data, pembersihan data, transformasi data, data mining, serta interpretasi atau evaluasi terhadap pola yang dihasilkan (Dewi et al., 2022).

2.4 Teknik Klasifikasi

Dalam proses klasifikasi, data dapat melalui beberapa tahapan, mulai dari ekstraksi atau pemilihan fitur, pengurangan dimensi apabila diperlukan, pemilihan algoritma klasifikasi, hingga evaluasi kinerja model. Tahapan tersebut bertujuan untuk memperoleh model klasifikasi yang mampu menghasilkan prediksi secara tepat sesuai dengan karakteristik data yang digunakan (Kowsari et al., 2019).

2.5 Data Set

Dataset merupakan sekumpulan data yang terorganisir dalam format tertentu, umumnya disusun dalam bentuk tabel yang terdiri atas baris dan kolom. Dataset digunakan sebagai sumber informasi dalam proses analisis dan penelitian serta menjadi dasar dalam pengembangan, pelatihan, evaluasi, dan validasi model *machine learning*. Oleh karena itu, dataset memiliki peran penting dalam proses pembelajaran mesin karena kualitas dan karakteristik data yang digunakan dapat memengaruhi hasil model yang dihasilkan (Alfayyadh et al., 2025).

2.6 Data Training

Dataset memiliki peranan penting dalam pengembangan model *machine learning* karena menjadi dasar dalam proses pelatihan, evaluasi, dan pengujian kemampuan model (Paullada et al., 2021).

2.7 Data Testing

Data testing merupakan sekumpulan data yang digunakan untuk menguji dan mengevaluasi kinerja model *machine learning* setelah proses pelatihan selesai. Data testing tidak digunakan dalam proses pelatihan model, sehingga dapat digunakan untuk mengetahui kemampuan model dalam menghasilkan prediksi terhadap data yang belum digunakan sebelumnya. Dengan demikian, penggunaan data testing dapat memberikan gambaran mengenai kemampuan model dalam melakukan klasifikasi atau prediksi pada data baru (Riccio et al., 2020).

2.8 Tahapan Pengolahan Data

Pengolahan data menggunakan tahapan Knowledge Discovery in Databases (KDD), yaitu:

1. Data Selection
2. Preprocessing
3. Transformation
4. Data Mining
5. Evaluation

Tahapan tersebut digunakan untuk memastikan data yang digunakan sesuai dengan kebutuhan proses klasifikasi.

2.9 Data Selection

Pada tahap data selection terdapat 178 data balita yang tersedia. Data dipilih berdasarkan relevansi dan kelengkapan atribut yang diperlukan untuk penelitian. Setelah proses seleksi dan dilanjutkan dengan pre processing, diperoleh 158 data yang digunakan pada tahap selanjutnya.

Tabel 1. Data Selection

No	Tanggal Lahir	Jenis Kelamin	TB	BB	LL	LK
1	27/02/2022	Laki-Laki	91	14	17	50
2	27/10/2021	Laki-Laki	99	17	17	50



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 8 Tahun 2026
ISSN 3025-0919 (media online)
Hal 2079-2086

3	30/04/2021	Laki-Laki	105	16	17	50
4	27/03/2022	Perempuan	89	11	15	46
5	22/06/2021	Perempuan	101	19	19	50
6	06/06/2022	Perempuan	89	11	14	47
7	02/12/2021	Laki-Laki	97	13	15	50
8	14/07/2021	Perempuan	97	14	16	48
9	01/01/2022	Laki-Laki	96	15	17	49
10	02/11/2020	Laki-Laki	105	17	17	49
...	...	...	...	...	...	...
178	20/03/2025	Laki-Laki	68,3	71	...	...

2.10 Preprocessing

Tahap preprocessing dilakukan untuk mempersiapkan data sebelum digunakan dalam proses klasifikasi. Tahapan ini meliputi pemeriksaan data yang hilang, penghapusan data yang tidak sesuai, pemeriksaan noise, dan penyusunan data agar lebih terstruktur.

Tabel 2. Data Preprocessing

No	S1	S2(BULAN)	S3 (KG)	S4 (TB/U)	S5(STA TUS)
1	Laki-Laki	41	14	91	stunting
2	Laki-Laki	45	17	99	normal
3	Laki-Laki	51	15,8	105	normal
4	Perempuan	40	11,1	89	stunting
5	Perempuan	49	18,5	101	normal
6	Perempuan	38	11	89	stunting
7	Laki-Laki	44	13,2	97	normal
8	Perempuan	48	13,9	96,9	normal
9	Laki-Laki	43	14,5	96	normal
10	Laki-Laki	57	17,2	105	normal
...	...	...	...	...	...
158	Laki-Laki	14	9,01	86,5	normal

Setelah preprocessing, diperoleh 158 data dengan atribut jenis kelamin, umur, berat badan, TB/U, dan status.

2.11 Transformation

Transformasi dilakukan agar data memiliki format yang sesuai dengan proses klasifikasi menggunakan Naive Bayes.

Tabel 3. Data Transformation

No	S1	S2	S3	S4	S5
1	L	41	14	91	stunting
2	L	45	17	99	normal
3	L	51	16	105	normal
4	P	40	11	89	stunting
5	P	49	18	101	normal
6	P	38	11	89	stunting
7	L	44	13	97	normal
8	P	48	14	96,9	normal
9	L	43	14	96	normal
10	L	57	17	105	normal
...	...	...	...	...	...



158	L	14	9	86,5	Normal
-----	---	----	---	------	--------

Jenis kelamin diubah menjadi dua kategori, yaitu L untuk laki-laki dan P untuk perempuan. Umur tetap menggunakan satuan bulan. Berat badan ditransformasikan menjadi bilangan bulat dengan pembulatan ke bawah. Variabel S4 menggunakan indikator TB/U. Status digunakan sebagai class atau label yang menjadi target klasifikasi.

3. ANALISA DAN PEMBAHASAN

3.1 Pembagian Data

Dataset yang telah melalui proses pengolahan kemudian dibagi menjadi data training dan data testing menggunakan metode sampling split pada RapidMiner. Rasio yang digunakan adalah 80:20.

Dari 158 data, diperoleh:

- 126 data training
- 32 data testing

Data training digunakan untuk membentuk model Naive Bayes, sedangkan data testing digunakan untuk menguji kemampuan model dalam melakukan klasifikasi terhadap data yang belum digunakan dalam proses pelatihan.

3.2 Naïve Bayes

Naive Bayes adalah algoritma pembelajaran sederhana yang memanfaatkan aturan Bayes bersama dengan kuat dengan asumsi bahwa atribut-atributnya independen secara kondisional berdasarkan kelasnya (Schröder & Schulz, 2022).

Teorema Bayes sendiri memberikan cara untuk menghitung probabilitas suatu kejadian berdasarkan pengetahuan sebelumnya tentang kondisi terkait. Rumus dasar Teorema Bayes adalah:

$$P(H|X) = \frac{P(X|H)p(H)}{p(x)}$$

Keterangan :

$P(H|X)P(H|X)P(H|X)$: Probabilitas hipotesis H berdasarkan data X (posterior)

$P(H)P(H)P(H)$: Probabilitas awal kelas H (prior)

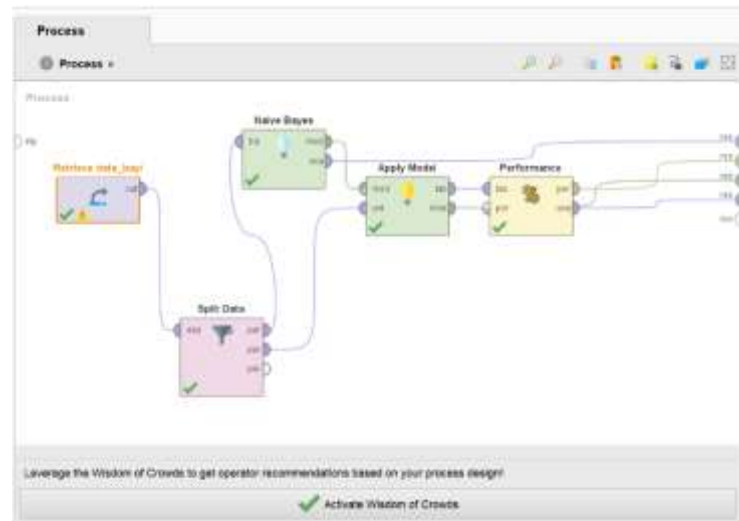
$P(X|H)P(X|H)P(X|H)$: Probabilitas data X pada kelas H

$P(X)P(X)P(X)$: Probabilitas data X

Dalam konteks klasifikasi, Naive Bayes menghitung probabilitas setiap kelas berdasarkan atribut-atribut input, kemudian memilih kelas dengan nilai probabilitas terbesar sebagai hasil klasifikasi.

3.3 Hasil Implementasi Rapid miner

Data yang telah melalui tahap selection, preprocessing, dan transformation kemudian dimasukkan ke dalam RapidMiner. Proses klasifikasi dilakukan dengan menyusun operator yang terdiri dari data training, data testing, Naive Bayes, Set Role, Apply Model, dan Performance Classification.

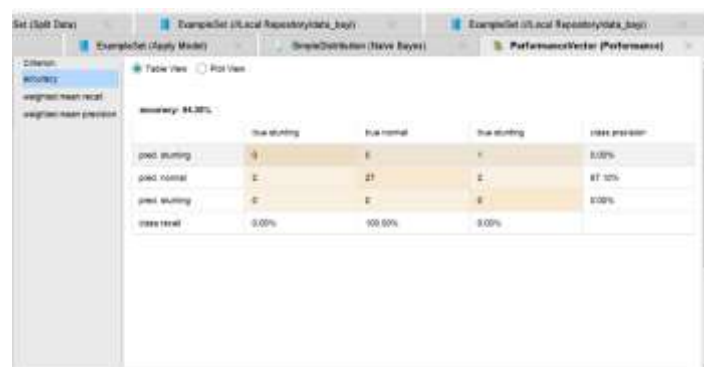


Gambar 1. Urutan Operator

Pada operator Performance Classification, parameter evaluasi yang digunakan adalah accuracy, mean recall, dan mean precision. Setelah seluruh operator terhubung, proses dijalankan untuk memperoleh hasil klasifikasi dan nilai performa model.

3.4 Hasil Pengujian Accuracy

Berdasarkan pengujian menggunakan 32 data testing, diperoleh 27 data yang diklasifikasikan dengan benar. Nilai accuracy dihitung sebagai berikut:



	True String	True Normal	True String	True Precision
pred string	4	1	1	0.00%
pred normal	1	27	1	87.10%
pred string	0	0	0	0.00%
True Recall	0.00%	100.00%	0.00%	

Gambar 2. Pengujian Accuracy

Accuracy = jumlah prediksi benar / jumlah seluruh data $\times 100\%$

Accuracy = $27 / 32 \times 100\%$

Accuracy = 84,38%

Dengan demikian, algoritma Naive Bayes memperoleh nilai accuracy sebesar **84,38%** pada data pengujian.

3.5 Hasil Pengujian Precision

Berdasarkan hasil pengujian pada RapidMiner, nilai mean precision yang diperoleh adalah 29,03%.



	true stunting	true normal	true stunting	class precision
pred. stunting	0	0	0	0.00%
pred. normal	0	27	2	87.10%
pred. stunting	0	0	0	0.00%
class recall	0.00%	100.00%	0.00%	

Gambar 3. Pengujian Precision

Pada kelas stunting, tidak terdapat data yang berhasil diprediksi dengan benar sebagai stunting sehingga precision kelas tersebut adalah 0%. Sementara itu, pada kelas normal terdapat 27 data yang berhasil diprediksi sebagai normal dengan precision sebesar 87,10%.

Hasil tersebut menunjukkan bahwa model lebih baik dalam menghasilkan prediksi kelas normal dibandingkan kelas stunting. Rendahnya nilai precision rata-rata menunjukkan bahwa model masih mengalami kesulitan dalam membedakan kelas yang jumlah datanya lebih sedikit.

3.6 Hasil Pengujian Recall

Hasil pengujian menunjukkan nilai mean recall sebesar 33,33%.



	true stunting	true normal	true stunting	class precision
pred. stunting	0	0	1	0.00%
pred. normal	0	27	2	87.10%
pred. stunting	0	0	0	0.00%
class recall	0.00%	100.00%	0.00%	

Gambar 4 Pengujian Recall

Pada kelas stunting, recall yang diperoleh adalah 0%. Hal tersebut disebabkan tidak terdapat data stunting yang berhasil diklasifikasikan dengan benar. Sebaliknya, kelas normal memperoleh recall sebesar 100% karena seluruh data normal yang digunakan pada pengujian berhasil dikenali sebagai normal.

Hasil ini menunjukkan adanya perbedaan kemampuan model dalam mengenali masing-masing kelas. Model memiliki kemampuan yang sangat baik dalam mengenali kelas normal, tetapi belum mampu mengenali kelas stunting secara optimal.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan pada implementasi algoritma Naive Bayes dalam pengklasifikasian kasus stunting pada Posyandu Mangga Jagakarsa, maka dapat ditarik kesimpulan sebagai berikut:

- Algoritma Naive Bayes berhasil diimplementasikan dengan baik dalam mengklasifikasikan status stunting balita berdasarkan variabel yang digunakan, yaitu jenis kelamin, umur (bulan), berat badan, dan nilai TB/U. Proses pengolahan data yang meliputi preprocessing,



transformasi data, pelatihan model, serta pengujian menggunakan tools RapidMiner dapat berjalan secara sistematis dan menghasilkan model klasifikasi yang stabil.

- b) Hasil pengujian model menunjukkan tingkat keberhasilan klasifikasi yang sangat tinggi, dimana nilai *accuracy* dan *weighted mean recall* yang diperoleh mencapai 84.38%. Hal ini menunjukkan bahwa algoritma Naive Bayes memiliki performa yang sangat baik dalam mengidentifikasi status stunting pada data balita di Posyandu Mangga Jagakarsa.

REFERENCES

- Adzim, F., et al. (2023). Klasifikasi status stunting balita menggunakan metode C4.5 berbasis web. 5(3), 515–525.
- Alfayyadh, M.B., Assegaff, S., & Fachruddin. (2025). Perbandingan algoritma C4.5 dan Naïve Bayes dalam. *Jurnal Manajemen Teknologi dan Sistem Informasi*, 5(2), 1095–1104.
- Dewi, S. P., Nurwati, N., & Rahayu, E. (2022). Penerapan data mining untuk prediksi penjualan produk terlaris menggunakan metode K-Nearest Neighbor. *Building Informatics, Technology and Science*, 3(4), 639–648. <https://doi.org/10.47065/bits.v3i4.1408>
- Irawati, S., Hermawati, E., Irawati, S., & Hermawati, E. (2023). Sensitive intervention policy recommendations to reduce stunting rates based on spatial analysis of sanitary factors in DKI Jakarta Province 2021. 8(3). <https://doi.org/10.7454/ihpa.v8i3.7158>
- Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information*, 10(4), 1–68. <https://doi.org/10.3390/info10040150>
- Lestari, E., Siregar, A., Hidayat, A. K., & Yusuf, A. A. (2024). Stunting and its association with education and cognitive outcomes in adulthood: A longitudinal study in Indonesia. *PloS one*, 19(5), e0295380. <https://doi.org/10.1371/journal.pone.0295380>
- Mulyono, M., Budianita, E., Nazir, A., & Syafria, F. (2023). Klasifikasi status stunting balita menggunakan metode Naïve Bayes Gaussian berbasis web. 8(3), 399–406.
- Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns*, 2(11), Article 100336. <https://doi.org/10.1016/j.patter.2021.100336>
- Riccio, V., Jahangirova, G., Stocco, A., Humbatova, N., Weiss, M., & Tonella, P. (2020). Testing machine learning based systems: A systematic mapping. *Empirical Software Engineering*, 25(6), 5193–5254. <https://doi.org/10.1007/s10664-020-09881-0>
- Schröder, T., & Schulz, M. (2022). Monitoring machine learning models: A categorization of challenges and methods. *Data Science and Management*, 5(3), 105–116. <https://doi.org/10.1016/j.dsm.2022.07.004>