

PREDIKSI PENDERITA DIABETES MENGGUNAKAN METODE NAIVE BAYES

Citra Agustina Rahayu^{1*}, Rudi Hartono², Aso Sudiarjo³

^{1,2,3}Teknik Informatika, Universitas Perjuangan; Jl. Peta No.177, Tasikmalaya, Jawa Barat

Riwayat artikel:

Received: 23 Mei 2023

Accepted: 10 Juli 2023

Published: 1 Agustus 2023

Keywords:

Prediksi; Diabetes Mellitus;
Data Mining; Naive Bayes.

Correspondent Email:

agustinarcitra@gmail.com

Abstrak. Prediksi adalah upaya memperkirakan sesuatu yang akan terjadi di masa depan dengan menggunakan berbagai informasi yang relevan pada waktu sebelumnya dengan menggunakan metode ilmiah. Data mining adalah kumpulan berbagai data atau informasi yang memiliki berbagai fungsi dan sangat bermanfaat di masa mendatang. Pemanfaatan data mining juga tidak hanya untuk teknologi saja melainkan bisa untuk Bidang Kesehatan salah satunya memprediksi penyakit diabetes dengan menggunakan metode Naive Bayes yaitu algoritma yang meminimalkan Kesehatan dari semua metode yang lain. Diabetes mellitus juga dikenal sebagai sekelompok masalah metabolisme yang ditandai dengan tingginya kadar gula darah yang berkepanjangan. Diabetes dapat menyebabkan masalah kesehatan yang serius termasuk penyakit kardiovaskular. Penelitian ini bertujuan untuk memprediksi dan mengukur tingkat akurasi penyakit Diabetes. Pengumpulan data yang dilakukan di RSUD dr.soekardjo dengan menghasilkan 502 data pada rekam medis July 2019 – Desember 2022. Pengolahan Naive Bayes menghasilkan prediksi dengan 5 kali pengujian yaitu sebesar 57,38%, 70,27%, 78,26%, 85,71% dan 88,93%. Akurasi tertinggi sebesar 95.92% artinya rule yang dihasilkan mendekati 100%.

Abstract. Prediction is an attempt to predict something that will happen in the future using various relevant information at a previous time using the scientific method. Data mining is a collection of various data or information that has various functions and is very useful in the future. The use of data mining is also not only for technology but can be for the Health Sector, one of which predicts diabetes using the Naive Bayes method, which is an algorithm that minimizes health from all other methods. Diabetes mellitus is also known as a group of metabolic problems characterized by prolonged high blood sugar levels. Diabetes can lead to serious health problems including cardiovascular disease. This study aims to predict and measure the accuracy of diabetes. Data collection was carried out at RSUD dr.soekardjo by producing 502 data in medical records July 2019 – December 2022. Naive Bayes processing produces predictions with 5 tests, namely 57.38%, 70.27%, 78.26%, and 85.71% and 88.93% the highest accuracy of 91.84% means that the resulting rule is close to 100%.

1. PENDAHULUAN

Prediksi (peramalan) adalah satu usaha untuk meramalkan peristiwa di masa depan melalui peristiwa masa lampau. Prediksi (peramalan) dilakukan dengan mengumpulkan sejumlah data sebelumnya (*history data*), di

mana data tersebut dianalisis dan dihitung untuk mendapatkan corak tertentu sehingga menghasilkan data prediksi (ramalan) untuk masa depan [1].

Data Mining adalah proses menemukan pola data dan informasi yang menarik dari kumpulan

data yang sangat besar. Sumber data dapat mencakup database, gudang data, jaringan, repositori, atau data yang dimasukkan ke dalam sistem dinamis [2].

Diabetes Mellitus juga dikenal sebagai (DM) adalah sekelompok masalah metabolisme yang ditandai dengan gula darah tinggi secara kronis. Gejala gula darah tinggi antara lain buang air kecil berlebihan, selalu merasa haus, lapar yang meningkat, berat badan turun, penglihatan kabur, dan luka yang sulit sembuh. Jika tidak diobati, diabetes dapat menyebabkan masalah kesehatan yang serius seperti *ketoasidosis* diabetis, *hiperglikemia*, *hyperosmolar*, dan bahkan kematian. Hal ini dapat menyebabkan komplikasi seumur hidup termasuk penyakit *kardiovaskular*, stoke, gagal ginjal, borok kaki, komplikasi mata dan lain-lain [3].

Berbagai faktor seperti genetika, gaya hidup tidak sehat, obesitas, dan usia dapat berkontribusi terhadap diabetes. Pada diabetes tipe 1, sistem kekebalan tubuh menyerang sel-sel yang memproduksi insulin. Pada diabetes tipe 2, tubuh menjadi kebal terhadap insulin atau tidak menghasilkan cukup insulin. Diabetes dapat menyebabkan komplikasi serius yang jika tidak diobati, akan menyebabkan kerusakan organ tubuh, *neuropati*, *retinopai*, dan masalah *kardiosvaskular*. Oleh karena itu sangat penting untuk menerapkan gaya hidup sehat dan menjalani pengobatan yang tepat untuk mencegah komplikasi serius [4].

Menurut Organisasi Kesehatan Dunia (*WHO*) Diabetes diabetes adalah salah satu penyakit kronis dengan tingkat kematian yang tumbuh paling cepat, yang mempengaruhi 422 juta orang di seluruh dunia pada tahun 2018 [5].

Menurut Kemenkes RI Diabetes Melitus di Jawa Barat naik dari 1,3% menjadi 1,7% dan menurut Dinas Kesehatan jumlah kasus diabetes melitus terus meningkat di Kota Tasikmalaya sampai akhir tahun 2020 berjumlah 4529 kasus, dimana setiap tahunnya mengalami peningkatan [6].

Menurut Rekam Medis di RSUD Dr. Soekardjo Tasikmalaya *Diabetes Mellitus* termasuk kedalam penyakit 10 besar penyakit terbanyak. Pada tahun 2017 terdapat 3.623 pasien rawat jalan dengan diabetes dan 352 pasien rawat inap yang sebagian besar adalah perempuan yaitu 2117 pasien rawat jalan, 212 pasien rawat inap, serta pasien diabetes pria

rawat jalan ada 1506, serta 140 pasien rawat inap termasuk dengan setiap tahun nya bertambah.

Algoritma *Naive Bayes* dapat meminimalisir kesehatan dibandingkan dengan semua algoritma lainnya. Oleh karena itu penulis mencoba mengaitkan kasus ini dengan metode *Naive Bayes* dengan 16 atribut ntuk mengetahui prediksi diabetes di rsud dr.soekardjo dan seberapa besar tingkat akurasi menggunakan metode tersebut. Metode ini pertama kali dikenalkan oleh Ilmuan Inggris yang Bernama Thomas Bayes. Yang dimana dapat memprediksi probabilitas dan statistic.

2. TINJAUAN PUSTAKA

2.1. Diabetes

Diabetes adalah kondisi kronis yang bermanifestasi sebagai jenis gangguan metabolisme dengan kadar gula darah yang meningkat. Jenis diabetes dikategorikan berdasarkan faktor-faktor yang berkontribusi terhadap peningkatan kadar gula darah [7].

2.2. Data Mining

Data mining merupakan sebuah proses yang dilakukan untuk melakukan pengolahan terhadap data yang besar yang berguna menemukan informasi – informasi baru tersimpan pada kumpulan data tersebut yang nantinya dapat dimanfaatkan dimasa yang akan datang [8].

2.3. Naive Bayes

Algoritma *Naive Bayes* menggunakan perhitungan probabilistik. Secara umum, teknik ini, yang terkenal dengan akurasi yang tinggi, sering digunakan untuk menyelesaikan tantangan prediksi berbasis klasifikasi [9]. *Naive Bayes* adalah algoritma klasifikasi populer yang digunakan dalam aplikasi pembelajaran mesin. Itu membuat asumsi bahwa setiap nilai atribut yang mempengaruhi nilai kelas adalah independen dari semua nilai atribut lain yang mempengaruhi nilai-nilai kelas lainnya. Metode *Naive Bayes* diciptakan menggunakan teorema Bayes, dan itu membuat asumsi bahwa setiap karakteristik independen dan unik dari kualitas lain sendiri [10].

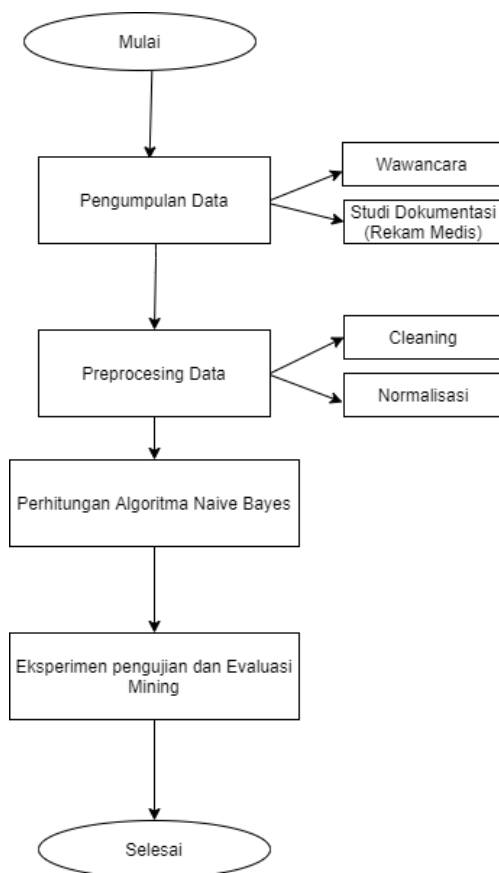
2.4. RapidMiner

RapidMiner adalah program *open source*. RapidMiner adalah mesin penambangan data yang dapat diintegrasikan ke dalam produknya sendiri dan tersedia sebagai perangkat lunak

mandiri untuk analisis data, untuk memberikan informasi kepada pengguna sehingga mereka dapat membuat keputusan terbaik [11].

3. METODE PENELITIAN

Metode dasar yang digunakan dalam penelitian ini adalah metode kuantitatif. Tujuannya yaitu melakukan prediksi dan evaluasi model algoritma *Naive Bayes* untuk mengetahui akurasi algoritma *Naive Bayes* dalam penyakit Diabetes.



Gambar 1 Metodologi Penelitian

3.1. Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan secara wawancara dan studi dokumentasi melalui rekam medis Rumah Sakit Umum Daerah dr.soekardjo pada July 2019 – Desember 2022, dengan mendapatkan 502 data. Data yang dikumpulkan diawali dengan melihat buku dan rekam medis elektronik. Setelah itu kemudian pendataan data pasien penderita *diabetes mellitus* dengan proses mengelompokan data ke dalam Microsoft excel.

3.2. Preprocessing Data

Teknik *preprocessing* ini digunakan dalam penelitian dengan pengumpulan data tujuannya agar proses mengubah data mentah menjadi format yang dapat dipahami. Seperti proses cleaning dan normalisasi. Normalisasi yang dilakukan menggunakan *min-max normalization* dengan rentang 0-1.

3.3. Perhitungan Algoritma Naive Bayes

Algoritma yang digunakan dalam penelitian ini menggunakan algoritma *Naive Bayes* dengan menggunakan Rapid Miner.

3.4. Eksperimen pengujian dan Evaluasi Mining

Eksperimen pengujian ini dengan split data. Evaluasi ini yaitu Teknik terhadap hasil akhir penelitian yang sudah dirancang dengan software yang sesuai dengan prosedur rancangan dengan menghitung akurasi dari kebenaran data dengan *confusion matrix*.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Hasil dan Diskusi menyusun 60-70% dari naskah. Bagian ini adalah bagian utama dari artikel penelitian. Hasil harus meringkas atau menyoroti temuan daripada memberikan hasil dan analisis rinci. Berisi hasil yang diambil dari analisis data dan/atau hasil uji hipotesis dan hanya menyediakan data yang mendukung pembahasan. Bagian ini meliputi tabel dan grafik yang diambil dari analisis data hasil penelitian.

Tabel 1 Data Mentah

No	U m ur	J K	B B	Sis tol e	Dia stol e	Glu kos a dar ah pua sa	...	Kel as
1	63	L	0	2	73	151	...	YA
2	64	L	5	0	68	95	...	TI DA K

3	64	P	5	11	74	97	...	TI
			5	2				DA
								K
4	62	L	6	15	63	103	...	TI
			4	2				DA
								K
5	52	P	1	13	86	222	...	TI
			5	3				DA
								K
...
5							...	TI
0			4					DA
2	57	P	1	86	58	96		K

Dalam penelitian ini memiliki 15 parameter yaitu umur, Jenis kelamin, berat badan, sistole, diastole, glukosa darah puasa, glukosa darah 2 jam pp, hba1c, trigliserida, kolesterol ldl, kolesterol hdl, kolesterol total, ureum, dan kreatin.

4.2. Preprocessing Data

Tabel 2 Hasil Normalisasi

N	U	J	BB	Sis	Dia	Glu	...	Ke
o	m	K		tol	stol	kos		las
ur				e	e	a		
						dar		
						ah		
						pua		
						sa		
1	63	L	0.3	0.3	0.1	0.2	...	YA
			13	78	64	40		
			8	5	7	3		
2	64	L	0.3	0.3	0.1	0.1	...	TI
			50	67	50	20		DA
			3	2	2	1		K
3	64	P	0.2	0.2	0.1	0.1	...	TI
			77	65	67	24		DA
			3	5	6	4		K
4	62	L	0.3	0.4	0.1	0.1	...	TI
			43	91	35	37		DA
			0	5	8	3		K
5	52	P	1.0	0.3	0.2	0.3	...	TI
				84	02	92		DA
				1	3	7		K
...

4	57	P	0.1	0.1	0.1	0.1	...	TI
8			75	18	21	22		DA
7			1	6	3	3		K

Rumus Normalisasi :

$$= \frac{\text{jumlah atribut ya dan tidak}}{\text{seluruh jumlah atribut kelas}}$$

Nilai minimum sistole : 65

Nilai minimal sistole : 242

Contoh perhitungan sistole

$$1. (132-65)/(242-65)=0,379$$

$$2. (130-65)/(242-65)=0,367$$

$$3. (112-65)/(242-65)=0,266$$

$$4. (152-65)/(242-65)=0,491$$

$$5. (133-65)/(242-65)=0,384$$

Contoh perhitungan normalisasi diastole:

$$1. (73-16)/(362-16)=0,164$$

$$2. (68-16)/(362-16)=0,150$$

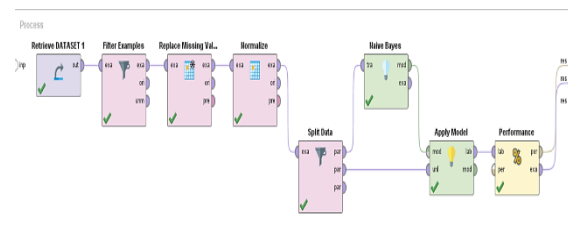
$$3. (74-16)/(362-16)=0,167$$

$$4. (63-16)/(362-16)=0,135$$

$$5. (86-16)/(362-16)=0,202$$

4.3. Perhitungan Algoritma Naive Bayes

Setelah data preprocessing data di atas tersebut maka langkah selanjutnya pengerjaan metode *Naive Bayes* pada dataset yang sudah ada pada RapidMiner.



Gambar 2 Lembar Kerja Pembagian Data menggunakan Split Data

Dimana operator *Naive Bayes* adalah algoritma *Naive Bayes*, sedangkan *Split Data* digunakan untuk membagi data latih dan data uji, *Apply Model* digunakan untuk mengeluarkan aturan dari *Naive Bayes*, dan *Performance operator* digunakan untuk mengeluarkan hasil atau akurasi model yang digunakan dengan akurat.

Perhitungan awal *Naive Bayes* yaitu dengan membagi seluruh atribut kelas dengan sebanyak jumlah data.

- Jumlah seluruh data : 487
- Kasus diabetes (1) : 399

c. Kasus non diabetes (2) : 88

$$= \frac{\text{jumlah atribut ya dan tidak}}{\text{seluruh jumlah atribut kelas}}$$

$$\text{atribut ya} = \frac{399}{487} = 0,819$$

$$\text{atribut tidak} = \frac{88}{487} = 0,181$$

4.4. Eksperimen pengujian dan Evaluasi Mining

Eksperimen pengujian menggunakan split data dengan 5 kali percobaan yaitu sebagai berikut :

Tabel 3 Confution Matrix

	true YA	true TIDAK	class precisio n
pred. YA	39 (TP)	1 (FN)	97.50%
pred. TIDAK	1 (FP)	8 (TN)	88.89%
class recall	97.50 %	88.89%	

$$\text{a) } Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Accuracy = \frac{39 + 8}{39 + 8 + 1 + 1} = 0.95918 = 95,92\%$$

$$\text{b) } Precision = \frac{TP}{TP+FN}$$

$$Precision = \frac{39}{39 + 1} = 0.975 = 97.50\%$$

$$\text{c) } NPV = \frac{TN}{TN+FP}$$

$$NPV = \frac{8}{8 + 1} = 0,88888 = 88,89\%$$

$$\text{d) } Recall = \frac{TP}{TP+FP}$$

$$Recall = \frac{39}{39 + 1} = 0,975 = 97,50\%$$

$$\text{e) } Specificity = \frac{TN}{TN+FN}$$

$$Specificity = \frac{8}{8 + 1} = 0,88888 = 88,89\%$$

Nilai precision, recall, dan accuracy untuk setiap eksperimen dapat ditentukan dari data pengujian. Angka precision, recall, dan accuracy yang diperoleh dari perhitungan rata-rata semua eksperimen adalah sebagai berikut:

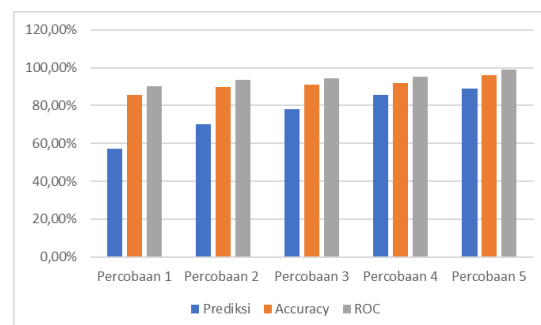
Tabel 4 Split Data Pengujian

Split Data	Accuracy
50% Data Training dan 50% Data Testing	85,60%
60% Data Trainining dan 40% Data Testing	89,74%
70% Data Training dan 30% Data Testing	91,10%
80% Data Training dan 20% Data Testing	91,84%
90% Data Training dan 10% Data Testing	95,92%

Algoritma pemodelan prediksi diabetes dengan metode *Naive Bayes* memprediksi dengan sangat baik, terlihat dari nilai akurasi terbesar dari semua percobaan, yaitu 90%, dan 10% pada tes kelima, yaitu 95,92% [12]. adalah 97,50% dan 92,86%. Dan untuk semua percobaan, nilai recall terbesar adalah 97,50% dan nilai recall terendah adalah 86,93%. Nilai precision terbesar dan terendah di semua percobaan masing-masing adalah 97,50% dan 92,86%. Dan untuk semua percobaan, nilai recall terbesar adalah 97,50% dan nilai recall terendah adalah 86,93%.

Tabel 5 Model Prediksi

Split Data	Prediksi	ROC
50% Data Training dan 50% Data Testing	57,38%	0,902
60% Data Trainining dan 40% Data Testing	70,27%	0,938
70% Data Training dan 30% Data Testing	78,26%	0,944
80% Data Training dan 20% Data Testing	85,71%	0,951
90% Data Training dan 10% Data Testing	88,89%	0,989



Gambar 3 Grafik Hasil Pengujian

Secara umum pembagian data menurut [13]. yaitu 90:10 sudah dikatakan paling bagus dan mendapat nilai akurasi 95,92%. Grafik ROC (*Reciver Operating Characteristic*) diatas menunjukkan algoritma *Naive Bayes* dari 5 percobaan split data yang memiliki nilai tertinggi AUC sebesar 0.989 yang artinya sangat bagus dalam hasil penelitian.

5. KESIMPULAN

Hasil analisis yang telah dibahas pada bab sebelumnya terkait Prediksi penderita diabetes menggunakan metode *Naive Bayes* pada Rumah Sakit Umum Daerah dr. Soekardjo dapat disimpulkan sebagai berikut :

1. Melakukan penerapan *Naive Bayes* dengan mengolah dataset Diabetes asli menjadi data praproses untuk keperluan perhitungan *Naive Bayes*. Proses ini melibatkan pembagian dataset menjadi data latih dan data uji dengan menggunakan alat bantu RapidMiner. Dalam melakukan diagnosis penyakit diabetes di Rumah Sakit Umum Daerah dr. Soekardjo, dengan metode *Naive Bayes* telah diuji dengan 5 kali percobaan yang menghasilkan prediksi sebesar 57,38%, 70,27%, 78,26%, 85,71%, dan 88,89%.
2. Akurasi yang dihasilkan mencapai 95,92% artinya rule yang dihasilkan mendekati 100% dengan hasil positif sebanyak 39 orang dan negatif 8 orang.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini. (*The author would like to thank the related parties who have provided support for this research.*)

DAFTAR PUSTAKA

- [1] L. Aryani, Fatmasari, Afriyudi, and N. Hadinata, "Prediksi jumlah siswa baru dengan menggunakan metode exponential smoothing (studi kasus: SMK Ethika Palembang)," *Bina Darma Conf. Comput.*, pp. 237–244, 2020, [Online]. Available: <https://conference.binadarma.ac.id/index.php/BDCCS/article/view/1615%0Ahttps://conference.binadarma.ac.id/index.php/BDCCS/article/download/1615/693>
- [2] U. M. D. E. C. D. E. Los, "No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title".
- [3] N. P. Tigga and S. Garg, "Prediction of Type 2 Diabetes using Machine Learning Classification Methods," *Procedia Comput. Sci.*, vol. 167, pp. 706–716, 2020, doi: 10.1016/j.procs.2020.03.336.
- [4] S. Soelistijo, "Pedoman Pengelolaan dan Pencegahan Diabetes Melitus Tipe 2 Dewasa di Indonesia 2021," *Glob. Initiat. Asthma*, p. 46, 2021, [Online]. Available: www.ginasthma.org.
- [5] W. Apriliah, I. Kurniawan, M. Baydhowi, and T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *Sistemasi*, vol. 10, no. 1, p. 163, 2021, doi: 10.32520/stmsi.v10i1.1129.
- [6] Fabiana Meijon Fadul, "濟無No Title No Title No Title," no. Dm, pp. 1–8, 2019.
- [7] Kementerian Kesehatan RI, "Infodatin tetap produktif, cegah, dan atasi Diabetes Melitus 2020," *Pusat Data dan Informasi Kementerian Kesehatan RI*, pp. 1–10, 2020.
- [8] Sharyanto and D. Lestari, "Penerapan Data Mining Untuk Menentukan Segmentasi Pelanggan Dengan Menggunakan Algoritma K-Means dan Model RFM Pada E-Commerce," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, pp. 866–871, 2022, doi: 10.30865/jurikom.v9i4.4525.
- [9] M. Y. Putra and D. I. Putri, "Pemanfaatan Algoritma *Naive Bayes* dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI," *J. Tekno Kompak*, vol. 16, no. 2, pp. 176–187, 2022.
- [10] R. Hayami, Soni, and I. Gunawan, "Klasifikasi Jamur Menggunakan Algoritma *Naive Bayes*," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 1, pp. 28–33, 2022, doi: 10.37859/coscitech.v3i1.3685.
- [11] D. Pascalina, R. Widhiastono, and C. Julianne, "Pengukuran Kesiapan Transformasi Digital Smart City Menggunakan Aplikasi Rapid Miner," *Technomedia J.*, vol. 7, no. 3, pp. 293–302, 2022, doi: 10.33050/tmj.v7i3.1914.
- [12] F. Gorunescu, "Data Mining: Concepts, Models, and Techniques," *Verlag Berlin Heidelberg: Springer*, 2011.
- [13] Rosebrock, "Deep Learning For Computer Vision," *United States Am. Pyimagesearch*, 2018.