
Prediksi Penyakit Stroke Pada Seseorang Menggunakan PySpark.ipynb

Fauzan Maulana
fauzanmaulana@gmail.com

Abstract — Stroke is the second deadliest disease in the world according to WHO. The sufferer has an injury to the nervous system. Because of this, health experts, especially in the field of nursing, need special attention. Currently the development of the Industrial Revolution Era 4.0 collaborates in the fields of technology and health sciences so that it becomes something useful by using Machine Learning. There are so many benefits used in predicting several diseases that can be anticipated. Stroke using the Adaboost algorithm which has advantages can be combined with the Naïve Bayes algorithm. The application of this method uses split data, namely training data and testing data which are made portions of the test. The first test results using Naive Bayes with 80/20 split data gives an accuracy of 0.976 which has a very good classification diagnosis. After optimization with Adaboost with a 70/30 split data it gives an accuracy of 0.981. From the previous test and after the optimization gave a difference of 0.005. The use of Adaboost as an optimization with Naïve Bayes can be used as a consideration to create a smart system that is used by health professionals for good decision making in the fields of nursing and medicine in accelerating the diagnosis of stroke patients.

Keywords — Stroke, Machine Learning, Naïve Bayes, Adaboost

Abstrak—Stroke adalah penyakit paling mematikan nomor dua di dunia menurut WHO. Penderitanya mengalami cedera pada sistem saraf. Karena hal inilah para pakar kesehatan khususnya dibidang keperawatan memerlukan perhatian khusus. Saat ini perkembangan Era Revolusi Industri 4.0 yang berkolaborasi di bidang teknologi dan ilmu kesehatan sehingga menjadi sesuatu yang bermanfaat dengan menggunakan Machine Learning. Banyak sekali manfaat yang digunakan dalam memprediksi beberapa penyakit yang dapat diantisipasi. Khususnya penyakit stroke dengan menggunakan algoritma Adaboost yang mempunyai kelebihan bisa digabung dengan algoritma Naïve Bayes. Penerapan metode ini menggunakan Split Data yaitu data training dan data testing dibuat porsi dalam melakukan pengujian. Hasil pengujian yang pertama menggunakan Naive Bayes dengan data split 80/20 memberikan akurasi 0,976 memiliki diagnosis klasifikasi yang sangat baik. Setelah di optimasi dengan Adaboost dengan data split 70/30 memberikan akurasi 0,981. Dari pengujian sebelumnya dan setelah di optimasi memberikan selisih sebesar 0,005. Penggunaan Adaboost sebagai optimasi dengan Naïve bayes dapat di gunakan sebagai bahan pertimbangan untuk membuat sistem pintar yang digunakan oleh para ahli kesehatan untuk pengambilan keputusan yang baik di bidang keperawatan dan

kedokteran dalam mempercepat hasil diagnosa pasien stroke.

Kata kunci—Stroke, Machine Learning, Naïve Bayes, Adaboost

I. INTRODUCTION

Stroke adalah penyebab utama kematian dan kecacatan secara global. Diagnosis tergantung pada fitur klinis dan pencitraan otak untuk membedakan antara stroke iskemik dan perdarahan intraserebral.[1][2] Stroke, atau kecelakaan serebrovaskular, melibatkan cedera pada sistem saraf pusat sebagai akibat dari penyebab vaskular, dan merupakan penyebab utama kecacatan di seluruh dunia.[3]. Stroke merupakan kematian tertinggi kedua di seluruh dunia dengan penyebab utama adalah kecacatan. Dari WHO sendiri 70% secara global disebabkan oleh stroke, 87% kematian akibat stroke terjadi di negara-negara berpenghasilan rendah dan menengah [4]. Stroke menjadi penyebab pertama kematian yang diprediksi pada tahun 2030 (14,4% dari total kematian) dan penyebab ketiga DALY lost (6% dari total DALY) di negara-negara berpenghasilan menengah; dan penyebab kematian yang ketiga (8,2% dari total kematian) dan 8 penyebab utama DALY lost (2,8% dari total DALY) di negara-negara berpenghasilan rendah[5]. Masalah penyakit stroke di Indonesia memerlukan perhatian yang serius karena jumlah kasus yang terus meningkat dan mempunyai angka kematian yang tinggi.[6]

Berdasarkan Laporan Hasil Riskesdas 2018 menunjukkan prevalensi stroke nasional yang masih cukup tinggi yaitu 10.9 per mil. Salah satu provinsi di Indonesia dengan prevalensi stroke yang tinggi adalah Sulawesi Selatan yaitu 10.6 per mil.

Di Indonesia sendiri dari diagnosis tenaga kesehatan untuk prevalensi stroke sebesar 7 per mil dan juga untuk gejala besarnya adalah 12,1 per mil. Menurut data dari tenaga kesehatan di Sulawesi Utara sebesar 10,8% kemudian di D.I Yogyakarta sendiri sebesar 10,3. Provinsi DKI Jakarta dan Bangka Belitung untuk prevalensi stroke masing-masing 9,7 per mil. Gejala stroke tertinggi juga berada di provinsi Sulawesi Selatan 17,9% di ikuti D.I Yogyakarta 16,9% kemudian Sulawesi Tengah 16,6%. Prevalensi Stroke di Jawa

Timur sebesar 16 per mil. Pada wilayah Kalimantan Selatan prevalensi penderita stroke yang terdiagnosis tenaga kesehatan sebanyak 9,2% dan yang terdiagnosis tenaga kesehatan dan gejala sebanyak 14,5%. [7] [8]

Prevalensi stroke juga terjadi di provinsi Kalimantan Selatan yang memiliki sebesar 9,7 per seribu jumlah penduduk dengan rentang rata-rata 5,2-18,5 per seribu penduduk.

Hasil analisa dari Dinas Kesehatan provinsi Kalimantan Selatan untuk prevalensi Stroke meningkat karena menyesuaikan peningkatan umur, dengan jenis kelasmin rata-rata di dominasi oleh perempuan, di ikuti dengan pendidikan yang rendah, dan tempat tinggal di pedesaan. Penyakit stroke juga sangat rentan dengan yang tidak bekerja. [9]

Prognosis penyakit stroke kebanyakan sembuh dengan kecacatan. Kalimantan selatan terutama Kota Banjarmasin angka kejadian stroke mencapai 4.031 kasus yang sembuh dengan gejala sisa. Gejala sisa ini menjadikan penderita stroke di Kota Banjarmasin memiliki kualitas hidup rendah, terlebih lagi dengan rendahnya dukungan keluarga.[10] Dalam pemanfaatan teknologi khususnya di bidang ilmu kesehatan dengan menggunakan pemodelan Machine Learning semakin disesuaikan karena dengan pembelajaran mesin dapat mempermudah dalam prediksi dalam melakukan penanganan penyakit.[11]

Sebagai contoh di dalam menangani pasien yang sering terlambat dalam melakukan tindakan, peluang pasien bisa diobati jika sebuah sistem penangan secara otomatis itu dilakukan, paling tidak dalam mengambil keputusan awal saat sebelum kejadian. Sehingga dibuat pemodelan pembelajaran mesin yang mampu bekerja secara otomatis agar digunakan secara langsung dalam membantu pasien untuk menekan resiko.[12] Penggunaan Machine Learning, terbukti telah banyak diterapkan dalam topik klasifikasi dan optimasi dalam membuat sistem cerdas untuk meningkatkan penyedia layanan kesehatan.[13][14]

Banyak cara telah dilakukan untuk memprediksi berbagai penyakit dengan membandingkan kinerja teknologi Data Mining prediktif. sebagai proses pemilihan fitur, algoritma analisis komponen prinsip digunakan untuk mengurangi dimensi dan mengadopsi algoritma klasifikasi dalam membangun model klasifikasi.[15] Klasifikasi adalah teknik untuk membentuk model data yang belum diklasifikasikan, maka model dapat digunakan untuk mengklasifikasikan data baru.[16] Dalam meningkatkan tingkat akurasi yang tinggi digunakan lah Optimasi. Pada penelitian ini untuk pengujian penyakit diabetes melitus menggunakan Naive Bayes Classifier dengan Particle Swarm Optimization, hasil yang diperoleh adalah metode Naive Bayes Classifier untuk 2 kelas dan 3 kelas masing-masing menghasilkan nilai akurasi 78,88% dan 68,50%. tetapi setelah menambahkan Particle Swarm Optimization nilai akurasi meningkat masing-masing menjadi 82,58% dan 71, 29%. Hasilnya adalah dari 2 kelas memiliki nilai akurasi lebih tinggi dari 3 kelas dengan perbedaan 11,29%.[17]

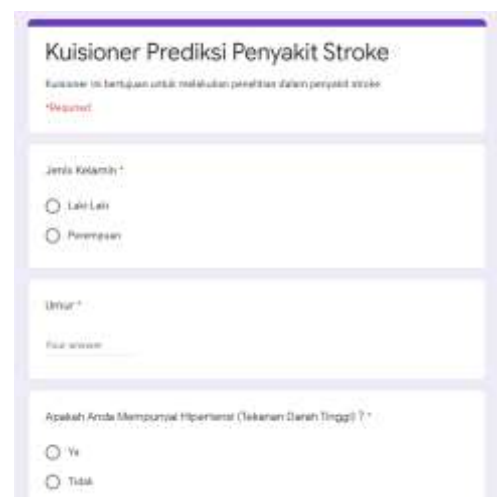
Penelitian berikutnya Berdasarkan hasil pengujian data pelatihan yang digunakan sebagai data pengujian untuk pasien penyakit stroke, total data 203, dapat dinyatakan bahwa algoritma Naive Bayes Classifier mencapai akurasi 89,65% dalam proses mengklasifikasikan data tetapi penelitian tersebut tanpa dioptimalkan.[18] Dalam pemilihan fitur digunakan untuk memilih fitur klasifikasi yang optimal untuk memastikan bahwa algoritma pendeteksian dapat dilatih secara efektif. Metode ini memanfaatkan algoritma Adaboost yang merupakan algoritma klasifikasi dalam bidang pembelajaran mesin. yang terlatih dapat secara efektif. [19]

Optimalisasi algoritma adaboost dilakukan dalam pemilihan fitur untuk mendiagnosa kanker paru-paru. Penggunaan memilih fitur klasifikasi yang optimal dalam memastikan bahwa algoritma pendeteksian dapat dilatih secara efektif. [20] Keakuratan klasifikasi dengan menggunakan algoritma adaboost mulai dari pengukuran F, dan area di bawah kurva karakteristik operasi penerima (AUC) digunakan untuk mengevaluasi kinerja metode pembelajaran mesin.[21]

Pada penelitian ini kami menggunakan algoritma adaboost dalam mengoptimasi dalam meningkatkan akurasi algoritma naïve bayes untuk prediksi penyakit stroke.

II. METODOLOGI PENELITIAN

Penelitian ini dilaksanakan di Rumah Sakit Daerah dan beberapa puskesmas di Banjarmasin, mengingat Kota Banjarmasin merupakan kota rujukan bagi rumah sakit daerah lainnya seperti KalTeng sehingga tingkat kejadian Stroke cukup meningkat. Metode penelitian yang digunakan mengikuti langkah-langkah untuk membuat model memprediksi stroke menggunakan algoritma naïve bayes dengan optimasi adaboost seperti pada Gambar 2.[22]Mulai dari latar belakang masalah hingga studi literatur. Pengumpulan data dilakukan dengan menyebar kuesioner menggunakan google form pada pasien yang di beberapa puskesmas di Banjarmasin dengan format seperti Gambar 1 [23];

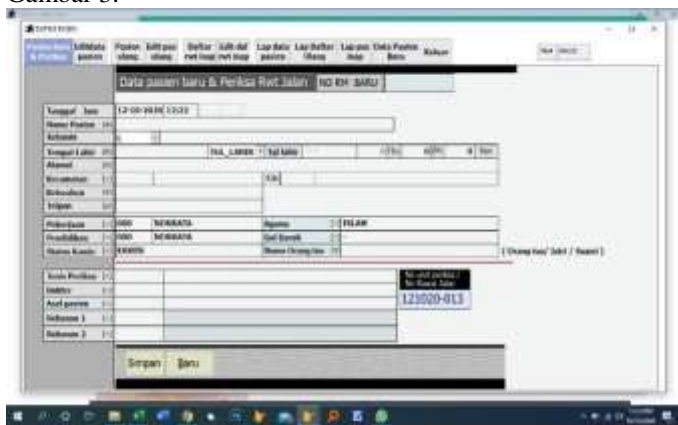


Gambar 1. Kuesioner Google Form



Gambar 2. Metode Penelitian

Dalam mengolah kuesioner peneliti mengadopsi dari variabel www.kaggle.com untuk data training dataset stroke. Data yang dikumpulkan dengan data testing diambil dari hasil dari kuisisioner dan sistem aplikasi rumah Sakit pada Gambar 3.



Gambar 3. Sistem Informasi Rumah Sakit

Data yang telah dikumpulkan kemudian di lakukan *cleaning data* untuk memasukan ke dalam 10 variabel(atribut) dan 1 variabel sebagai label target yaitu stroke dapat dilihat pada Tabel I.

TABEL I. VARIABEL DATASET STROKE

No	Variabel	Deskripsi
1	Jenis Kelamin	Laki-laki/Perempuan
2	Umur	Umur (dalam angka)
3	Hipertensi	Iya/tidak
4	Serangan Jantung	Iya/tidak
5	Status Pernikahan	Iya/tidak
6	Pekerjaan	Pedagang/mandiri/pns/pengangguran
7	Tempat tinggal	Kota/desa
8	Rata-rata gula darah	Ukuran gula darah (angka)
9	BMI	Bodi mass index (angka)
10	Status Merokok	Merokok/tidak/kadang-kadang
11	Stroke	Iya/tidak (target)

Data yang terkumpul melalui kuesioner dipisah menjadi data testing dengan jumlah data 572 baris. Data training sebanyak 28.500. Tahap pengujian menggunakan split data yaitu membagi data menjadi 2 untuk melakukan pengujian model algoritma naïve bayes dengan optimasi adaboost.[22] Dalam tahap pengujian peneliti menggunakan tools/software bahasa pemrograman Python 3.9 dengan tambahan plugin; pandas, numpy, scikit learn, matplotlib, dan jupyter notebook.[24]

III. HASIL DAN PEMBAHASAN

A. Hasil Pengujian Naïve Bayes

Na ıve Bayes adalah algoritma pembelajaran sederhana yang memanfaatkan aturan Bayes bersama-sama dengan kuat sebagai-sumption bahwa atribut kondisional di-tergantung diberikan kelas. Kelebihan yang dimiliki Bayesian classification adalah dapat memprediksi probabilitas didalam suatu class keanggotaan, karena sangat serupa dalam klasifikasi dari algortima decision tree dan neural network.[25] Jumlah data yang besar dapat di terapkan dengan Bayesian classification karena sangat terbukti mempercepat hasil akurasi yang terbaik.[16] hasil pengujian menggunakan splitting data yang menghasilkan hasil prediksi dari Naïve Bayes berikut TABEL II hasil;

TABEL II HASIL PENGUJIAN NAİVE BAYES

No	Split Data	Akurasi
1	60-40	0.974
2	70-30	0.975
3	80-20	0.976
4	90-10	0.972

Hasil dari Table I tersebut terlihat hasil splitting data yang paling tinggi adalah 80% /20 % dengan nilai Akurasi sekitar 0.976. Dibawah ini adalah hasil eksekusi menggunakan *tools*

Prediksi Algoritma Naive Bayes dengan split Data 80%/20% :

```

In [4]: #train test split
array = df.stroke.values
X = array[:,0:10]
y = array[:,10]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)

In [5]: #model
cnb = CatBoostClassifier()
model_cnb = cnb.fit(X_train, y_train)
y_pred_cnb = cnb.predict(X_test)

In [6]: print("Hasil Dari Confusion Matrix :")
print(metrics.confusion_matrix(y_test, y_pred_cnb))
print("Akurasi:", metrics.accuracy_score(y_test, y_pred_cnb))

Hasil Dari Confusion Matrix :
[[5672  46]
 [ 93   4]]
Akurasi: 0.9760963026655202

```

Gambar 4. Hasil Eksekusi dari tools python Jupyter notebook

Hasil Dari Confusion Matrix :
[[5672 46]
[93 4]]
Akurasi: 0.9760963026655202

Berikut hasil data Confusion Matrix, sebagai berikut hasil dari eksekusi python pada Gambar 4;

Pada Tabel II bahwa untuk jumlah True Positive (TP) adalah 5672, untuk True Negatif (TN) adalah 46, untuk False Positif (FP) adalah 93, dan untuk False Negative (TN) adalah 4.

B. Hasil Pengujian Naïve Bayes dengan Adaboost

Optimasi menggunakan adaboost dengan algoritma naïve bayes. Adaboost algoritma ensemble boosting sederhana melalui proses iterasi. Classifier-classifier yang digabungkan disebut weak classifier.[26][27] Salah satunya digunakan umumnya adalah classifier yang sederhana seperti algoritma naïve bayes. hasil pengujian menggunakan splitting data yang menghasilkan hasil prediksi dari adaboost dengan Naïve Bayes berikut table hasil;

TABEL III HASIL PENGUJIAN ADABOOST+NAÏVE BAYES

No	Split Data	Akurasi
1	60-40	0.9811
2	70-30	0.9816
3	80-20	0.979
4	90-10	0.978

Hasil dari Table III tersebut terlihat hasil splitting data yang paling tinggi adalah 70% /30 % dengan nilai Akurasi sekitar 0.9816. Berikut Tabel hasil data Confusion Matrix, sebagai berikut ;

Hasil Dari Confusion Matrix :
[[8560 5]
[155 2]]
Accuracy: 0.9816555835817473

Pada hasil eksekusi di atas bahwa untuk jumlah True Positive (TP) adalah 8560 , untuk True Negatif (TN) adalah 155, untuk False Positif (FP) adalah 5, dan untuk False Negative (TN) adalah 2.

C. Analisa Pembahasan

Dari hasil pengujian diatas, dengan dilakukan evaluasi baik secara confusion matrix beberapa algoritma yang diujikan terlihat hasil yang memiliki akurasi paling tinggi adalah optimasi adaboost dengan naïve bayes akurasi sebesar 0,981 ini lebih besar dari algoritma naïve bayes yang belum di optimasi ada selisih perbedaan sebelum di optimasi dan sesudahnya. Selisihnya adalah 0,005 berikut ini hasil perbedaan tersebut dapat dilihat pada Tabel IV dibawah ini;

TABEL IV HASIL PERBEDAAN AKURASI

Split Data	Model	Akurasi
80-20	Naïve Bayes	0.976
70-30	Adaboost + Naïve Bayes	0.981
Selisih		0.005

Dari hasil Tabel IV di atas kedua model tersebut sama-sama memiliki diagnose Excellent Classification, tetapi ada perbedaan pada saat pengujian split data, algoritma Naïve bayes menggunakan 80/20 dan 70/30 untuk naïve bayes + adaboost yang menghasilkan nilai akurasi maksimal.

IV. PENUTUP

Hasil penelitian untuk nilai akurasi algoritma Naïve Bayes memiliki nilai 0.976 dengan Split data 80/20, sedangkan untuk nilai akurasi optimasi adaboost dengan naïve bayes senilai 0.981 split data 70/30 kedua model tersebut memiliki diagnosa Excellent Classification, dalam pengujian prediksi penyakit stroke dengan dengan 11 variabel dengan jumlah data 28,500 untuk data training dan 572 untuk data testing (hasil dari kuisioner dan aplikasi di rumah sakit). Algoritma adaboost mempunyai kelebihan yaitu meningkatkan mengoptimasi yang bisa digabung dengan algoritma naïve bayes sebagai algoritma estimator sehingga menghasilkan akurasi yang dapat meningkatkan hasil yang terbaik dengan memiliki margin error yang kecil.

Untuk penelitian kedepannya akan dibuat sebuah aplikasi atau smart system application untuk memprediksi penyakit stroke, dengan menambah model dan metode lain yang ada di machine learning.

ACKNOWLEDGEMENT

Sebagai Peneliti Mengucapkan Terima kasih kepada Ristek-Brin yang telah membiayai penelitian ini, Dengan Skema Penelitian Dosen Pemula.

REFERENCES

- [1] B. C. V Campbell and P. Khatri, "Stroke," *Lancet*, vol. 396, no. 10244, pp. 129–142, Jul. 2020, doi: 10.1016/S0140-6736(20)31179-X.
- [2] G. J. Hankey, "Stroke," *The Lancet*. 2017, doi: 10.1016/S0140-

- [3] T. Ouyang and S. Sabat, "Stroke," in *Radiology Fundamentals: Introduction to Imaging & Technology*, 2020.
- [4] M. S. Phipps and C. A. Cronin, "Management of acute ischemic stroke," *The BMJ*. 2020, doi: 10.1136/bmj.l6983.
- [5] World Health Organization, "WHO - The top 10 causes of death," 24 Maggio, 2018. .
- [6] D. W. Nugraha, A. Y. E. Dodu, and N. Chandra, "Klasifikasi Penyakit Stroke Menggunakan Metode Naive Bayes Classifier (Studi Kasus Pada Rumah Sakit Umum Daerah Undata Palu)," *semanTIK*, vol. 3, no. 2, pp. 13–22, 2017.
- [7] Riskesdas, "Riset Kesehatan Dasar 2018," Kementrian Kesehat. Republik Indones., 2018, doi: 1 Desember 2013.
- [8] K. Riskesdas, "Hasil Utama Riset Kesehata Dasar (RISKESDAS)," *J. Phys. A Math. Theor.*, 2018, doi: 10.1088/1751-8113/44/8/085201.
- [9] Dinkes Kalsel, "Dinas Provinsi Kalimantan Selatan-Laporan Kinerja Bidang Kesehatan Masyarakat Tahunn 2018," *J. Chem. Inf. Model.*, 2018.
- [10] Oktovin, N. Elly, and M. Syafwani, "Studi fenomenologi pengalaman keluarga suku banjar selama merawat anggota keluarga dengan kondisi stroke di banjarmasin," *Keperawatan Suaka Insa.*, 2020.
- [11] J. Heo, J. G. Yoon, H. Park, Y. D. Kim, H. S. Nam, and J. H. Heo, "Machine Learning-Based Model for Prediction of Outcomes in Acute Stroke," *Stroke*, vol. 50, no. 5, pp. 1263–1265, May 2019, doi: 10.1161/STROKEAHA.118.024293.
- [12] V. Adelina and R. et All, "Klasifikasi Tingkat Risiko Penyakit Stroke Menggunakan Metode GA-Fuzzy Klasifikasi Tingkat Risiko Penyakit Stroke Menggunakan Metode GA- Fuzzy Tsukamoto," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. September, pp. 3015–3021, 2018.
- [13] I. Cholissodin, F. Farisuddin, and E. Santoso, "Klasifikasi Tingkat Resiko Stroke Menggunakan Improved Particle Swarm Optimization dan Support Vector Machine," *Konf. Nas. Sist. Inf. 2016, STT Ibnu Sina*, no. February, pp. 11–13, 2017.
- [14] K. Shameer, K. W. Johnson, B. S. Glicksberg, J. T. Dudley, and P. P. Sengupta, "Machine learning in cardiovascular medicine: are we there yet?," *Heart*, vol. 104, no. 14, pp. 1156–1164, Jul. 2018, doi: 10.1136/heartjnl-2017-311198.
- [15] M. S. Singh and P. Choudhary, "Stroke prediction using artificial intelligence," in *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, 2017, pp. 158–161, doi: 10.1109/IEMECON.2017.8079581.
- [16] B. Letham, C. Rudin, T. H. McCormick, and D. Madigan, "Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model," *Ann. Appl. Stat.*, 2015, doi: 10.1214/15-AOAS848.
- [17] D. S. Susilawati and D. Riana, "Optimization the Naive Bayes Classifier Method to diagnose diabetes Mellitus," *IAIC Trans. Sustain. Digit. Innov.*, vol. 1, no. 1, pp. 78–86, Nov. 2019, doi: 10.34306/itsdi.v1i1.21.
- [18] K. Vembandasamy, R. Sasipriya, and E. Deepa, "Heart Diseases Detection Using Naive Bayes Algorithm," *Int. J. Innov. Sci. Eng. Technol.*, 2015.
- [19] D. Tang, L. Tang, R. Dai, J. Chen, X. Li, and J. J. P. C. Rodrigues, "MF-Adaboost: LDoS attack detection based on multi-features and improved Adaboost," *Futur. Gener. Comput. Syst.*, 2020, doi: 10.1016/j.future.2019.12.034.
- [20] P. M. Shakeel, A. Tolba, Z. Al-Makhadmeh, and M. M. Jaber, "Automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks," *Neural Comput. Appl.*, 2020, doi: 10.1007/s00521-018-03972-2.
- [21] R. Garg, E. Oh, A. Naidech, K. Kording, and S. Prabhakaran, "Automating Ischemic Stroke Subtype Classification Using Machine Learning and Natural Language Processing," *J. Stroke Cerebrovasc. Dis.*, 2019, doi: 10.1016/j.jstrokecerebrovasdis.2019.02.004.
- [22] M. J. Zaki and W. Meira, Jr, *Data Mining and Machine Learning*. 2020.
- [23] B. Febriadi and N. Nasution, "Sosialisasi Dan Pelatihan Aplikasi Google Form Sebagai Kuisioner Online Untuk Meningkatkan Kualitas Pelayanan," *INOVTEK Polbeng - Seri Inform.*, 2017, doi: 10.35314/isi.v2i1.119.
- [24] J. Cutler and M. Dickenson, "Introduction to Machine Learning with Python," 2020.
- [25] D. Berrar, "Bayes' Theorem and Naive Bayes Classifier," in *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, 2019, pp. 403–412.
- [26] M. Hambali, Y. Saheed, T. Oladele, and M. Gbolagade, "ADABOOST Ensemble Algorithms for Breast Cancer Classification," *J. Adv. Comput. Res.*, 2019.
- [27] Q. Huang, Y. Chen, L. Liu, D. Tao, and X. Li, "On Combining Biclustering Mining and AdaBoost for Breast Tumor Classification," *IEEE Trans. Knowl. Data Eng.*, 2020, doi: 10.1109/TKDE.2019.2891622.