Volume XXI No. 1 Maret 2019
P-ISSN 1410-5063, E-ISSN: 2579-3500
Akreditasi Ristekdikti, No: 21/E/KPT/2018
http://ejournal.bsi.ac.id/ejurnal/index.php/paradigma

PARADIGMA
Jurnal Komputer dan Informatika Akademi Bina Saran Informatika

DOI: 10.31294/p.v20i2

Implementasi Algoritma *Naïve Bayes* Pada Data Set Hepatitis Menggunakan *Rapid Miner*

Deny Novianti

STMIK Nusa Mandiri Jakarta
e-mail: ddenynovianti@gmail.com

Cara Sitasi: Novianti, D. (2019, Maret). Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner. (S. Dalis, Ed.) Paradigma - Jurnal Komputer dan Informatika, 21(1). doi:10.31294/p.v21i1.4979

Abstract *- One of the diseases anticipated by doctors is hepatitis. This is because if the patient is not detected from the beginning having hepatitis, then the disease will develop into liver cancer. It can be seen that cancer is one of the deadliest diseases in the world that there are no drugs used for healing. By utilizing this increasingly developing science, researchers try to predict or classify whether a patient has suffered from hepatitis sickness based on the results of tests that have been undertaken before. One data mining technique can be used to predict Hepatitis and the method used is Naive Bayes. The data used is sourced from the UCI Repository with the web address https://archive.ics.uci.edu/ml/datasets/Hepatitis . The amount of data available is 155 data with 123 patients with Life hepatitis and 32 patients with Die hepatitis. The attributes contained in this hepatitis dataset are: Age, Sex, Steroids, Antivirals, Fatigue, Malaise, Anorexia, Big Liver, Liver Firm, Spleen Palpable, Spiders, Ascites, Varices, Bilirubin, Alk Phosphate, Sgot, Albumin, Protime, Histology , and Class (predictive result attribute). From the results of the research that has been done, it can be concluded that the Naive Bayes method includes an accurate algorithm to predict because the results of accuracy using Rapid Miner show more than 50% which is equal to 76.77%. With the highest Precision Class results of 98.88% for "Life" predictions, and Class Recall of 96.88% for "Die" Predictions.*

Keywords: Algorithms, Data Mining, Hepatitis, Naive Bayes, Rapid Miner

PENDAHULUAN

Dengan berkembangnya teknologi yang semakin pesat pada saat ini, menjadikan banyak hal yang sebelumnya sulit untuk diprediksi menjadi mudah untuk diprediksi. Dengan memanfaatkan ilmu pengetahuan yang ada dan didukung oleh data yang sesuai, maka untuk memprediksi suatu hal bukan lagi menjadi hal yang sulit. Di bidang pemerintahan, pendidikan, dan juga kesehatan kini sudah mulai memanfaatkan teknologi yang semakin berkembang ini. Terutama dalam bidang kesehatan, yang digunakan untuk mengklasifikasikan apakah sang pasien terdiagnosa dengan benar memiliki sebuah penyakit berdasarkan hasil tes yang telah dijalani.

Beberapa tahun ini, penyakit yang di antisipasi oleh para dokter salah satunya adalah penyakit Hepatitis. Hal tersebut disebabkan jika sang pasien tidak terdeteksi sejak awal memiliki penyakit hepatitis, maka penyakit tersebut akan berkembang menjadi penyakit kanker hati. Dapat diketahui bahwa kanker merupakan salah satu penyakit mematikan di dunia yang belum ada obat yang digunakan untuk penyembuhannya. Penyakit Hepatitis merupakan masalah kesehatan masyarakat di dunia termasuk di Indonesia, yang terdiri dari Hepatitis A, B, C, D, dan E. Hepatitis A dan E sering muncul sebagai kejadian luar biasa, ditularkan secara *fecal oral* dan biasanya berhubungan dengan perilaku *Life* bersih dan sehat, bersifat akurat dan dapat sembuh dengan baik. Sedangkan Hepatitis B, C, dan D (jarang) ditularkan secara parenteral, dapat menjadi kronis dan menimbulkan *cirrhosis* dan lalu kanker hati. Virus Hepatitis B telah menginfeksi sejumlah 2 milyar orang di dunia, sekitar 240 juta orang diantaranya menjadi pengidap Hepatitis B kronik, sedangkan untuk penderita Hepatitis C di dunia diperkirakan sebesar 170 juta orang. Sebanyak 1,5 juta penduduk dunia meninggal setiap tahunnya karena Hepatitis (Kementrian Kesehatan RI : 2014, n.d.)

Dengan memanfaatkan ilmu pengetahuan yang semakin berkembang ini lah para peneliti berusaha untuk memprediksi atau mengklasifikasikan apakah seorang pasien telah menderita sakit hepatitis berdasarkan hasil tes yang telah dijalani sebelumnya. Beberapa hasil penelitian terdahulu dapat disimpulkan bahwa algoritma Naive Bayes dapat digunakan untuk meneliti prediksi penyakit Hepatitis. Sistem diagnosis penyakit hati menggunakan metode Naïve Bayes dapat

diimplementasikan dengan 3 proses utama yaitu menghitung nilai prior atau peluang penyakit, menghitung likelihood berdasarkan masukkan pengguna, serta menghitung posterior yang diperoleh dari perkalian antara prior dan likelihood. Nilai posterior tertinggi akan diambil sebagai keputusan akhir sistem. Pengujian akurasi memperoleh hasil akurasi sebesar 87,5% dari 40 data uji terdapat lima ketidakcocokan antara hasil sistem dengan hasil diagnosis dokter, ketidakcocokan terjadi disebabkan karena gejala dimiliki oleh dua penyakit sedangkan sistem hanya dapat mendiagnosis dengan output satu penyakit (Prayoga, Hidayat, & Dewi, 2018). Metode naive Bayes dapat digunakan dalam memprediksi risiko seseorang terkena penyakit jantung. Ketepatan hasil prediksi terhadap hasil pengklasifikasian risiko penyakit jantung berdasarkan data yang didapatkan dari RSUD AWS bulan November dan Desember 2016 menggunakan program Delphi 7 Enterprise yaitu untuk percobaan 1 dengan jumlah data testing sebanyak 25 data didapat tingkat akurasi sebesar 80% dan pada percobaan 2 dengan jumlah data testing sebanyak 50 data diperoleh tingkat akurasi sebesar 78% (Sabransyah, Nasution, & Tisna, 2017). Didapatkan bahwa ketepatan klasifikasi menggunakan model *Naive Bayes* pada penelitian ini yaitu sebesar 93% atau memiliki *error* sebesar 7% (Ria Amora dan Akhmad Fauzy. 2016). Maka dari itu disimpulkan bahwa metode *Naive* Bayes dapat digunakan dalam melakukan prediksi untuk penyakit hepatitis. Dalam membangun sistem diagnosis penyakit kambing, metode yang digunakan adalah Naive Bayes. Proses dimulai dengan menerima inputan fakta kemudian dihitung dengan metode Naive Bayes. Akurasi pengujian sistem implementasi metode Naive Bayes untuk diagnosis penyakit kambing sebesar 90%. Ini dikarenakan metode Naive Bayes hanya melakukan perhitungan berdasarkan data latih dan inputan dalam sistem yang dibuat hanya berjumlah 4 inputan (Prayoga et al., 2018). Untuk mengimplementasikan metode Naïve Bayes data-data harus terlebih dahulu dikonversi menjadi bentuk diskrit dengan cara dikelompok-kelompokkan. Dengan menerapkan metode Naïve Bayes pada aplikasi prediksi maka didapatlah tingkat akurasi sebesar 82,97% (Bari, Sitorus, & Ristian, 2018).

1. Hepatitis

Istilah "Hepatitis" dipakai untuk semua jenis peradangan pada sel-sel hati, yang bisa disebabkan oleh infeksi (virus, bakteri, parasit), obat-obatan (termasuk obat tradisional), konsumsi alkohol, lemak yang berlebih dan penyakit autoimmune. Ada 5 jenis Hepatitis Virus yaitu Hepatitis A, B, C, D, dan E. Antara Hepatitis yang satu dengan yang lain tidak saling berhubungan. Hepatitis virus merupakan sebuah fenomena gunung es, dimana penderita yang tercatat atau yang datang ke layanan kesehatan lebih sedikit dari jumlah penderita sesungguhnya. Mengingat penyakit ini adalah penyakit kronis yang menahun, dimana pada saat orang tersebut telah terinfeksi, kondisi masih sehat dan belum menunjukkan gejala dan tanda yang khas, tetapi penularannya terus berjalan (Kementrian Kesehatan RI : 2014, n.d.)

2. Diagnosis

Diagnosis adalah suatu analisis terhadap kelainan atau salah penyesuaian dari pola gejala-gejalanya. Sama dengan istilah dalam dunia kedokteran, diagnosis merupakan kegiatan untuk menentukan jenis penyakit dengan meneliti gejala-gejalanya. Berdasarkan hal tersebut diagnosis merupakan proses pemeriksaan terhadap hal-hal yang dianggap tidak beres atau bermasalah (Suryanih, 2011).

Menurut (Kajian Pustaka, n.d.), diagnosis dapat diartikan sebagai:

a. Upaya atau proses menemukan kelemahan atau penyakit (weakness, disease) apa yang dialami seorang dengan melalui pengujian dan studi yang saksama mengenai gejala-gejalanya (symptons)

b. Studi yang saksama terhadap fakta tentang suatu hal untuk menemukan karakteristik atau kesalahan dan sebagainya yang esensial

c.Keputusan yang dicapai setelah dilakukan suatu studi yang saksama atas gejala-gejala atau fakta tentang suatu hal.

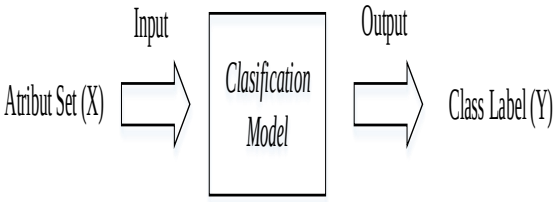
3. Data *Mining*

Data *mining*, sering juga disebut *Knowledge Discovery In Database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data *historis* untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Keluaran dari data *mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Saat ini istilah pengenalan pola (*pattern recognition*) jarang digunakan karena ia termasuk bagian dari data mining (Prasetyo, 2012).

Berdasarkan fungsinya, data mining dikelompokkan menjadi 6 yaitu deskripsi, estimasi, prediksi, klasifikasi, clustering, danasosiasi (Larose, 2005). Klasifikasi (taksonomi) adalah proses menempatkan objek tertentu (konsep) dalam satu set kategori, berdasarkan masing-masing objek (konsep) *property* (Gorunescu, 2011). Proses klasifikasi didasarkan pada empat komponen mendasar yaitu kelas, prediktor, training set, dan pengujian dataset.
Diantara model klasifikasi yang paling popular adalah *Decision/Classification Trees, Bayesian Classifiers/Naïve Bayes Classifiers, Neural Networks, Statistical Analysis, Genetic Algorithms, Rough Sets, K-Nearest Neighbor Classifier, Rule-based Methods, Memory Based Reasoning, Support Vector Machines* (Gorunescu, 2011).

4. Metode Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data ke dalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Model itu sendiri bisa berupa aturan "jika-maka", berupa pohon keputusan, atau formula *matedies* (Bustami, 2014).



Gambar 1. Diagram Model Klasifikasi

5. Metode Naive Bayes

Teori keputusan *bayes* adalah pendekatan statistik yang fundamental dalam pengenalan pola *(pattern recognition*) , pendekatan ini didasarakan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas dan ongkos yang di timbulkan dalam keputusan tersebut [9]. Selain itu *Bayes*ian *clasification* juga dapat memprediksi probabilitas keanggotaan suatu *Class* pada teorema *bayes* yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* . *Bayesian Classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar (Jananto, 2013).

Teorema *Bayes* memiliki bentuk umum sebagai berikut :

$$P(H|X) = \frac{P(X|H)\,P(H)}{P(X)} \quad \text{.....(11)}$$

Keterangan :

X = Data dengan *Class* yang belum diketahui
H = Hipotesis data X merupakan suatu *Class* spesifik
P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)
P(H) = Probabilitas hipotesis H (prior prob.)
P(X|H) = Probabilitas X berdasarkan kondisi tersebut
P(X) = Probabilitas dari X

6. *Rapid Miner*

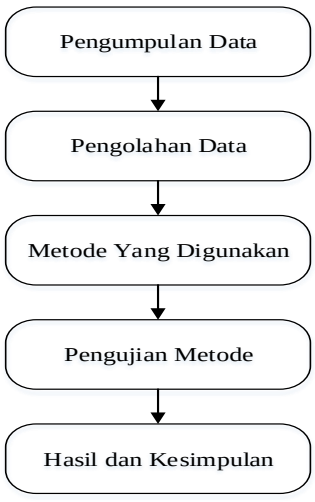
Rapid Miner adalah sebuah *tools* yang digunakan dalam teknik yang berada di lingkungan machine learning, data mining, *text mining* dan *predictive analytics* (Muslehatin & Ibnu, 2017).

Rapid Miner merupakan software/perangkat lunak untuk pengolahan data. Dengan menggunakan prinsip dan algoritma data mining, RapidMiner mengekstrak pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database. Rapid Miner memudahkan penggunanya dalam melakukan perhitungan data yang sangat banyak dengan menggunakan operator-operator. Operator ini berfungsi untuk memodifikasi data. Data dihubungkan dengan node-node pada operator kemudian kita hanya tinggal menghubungkannya ke node hasil untuk melihat hasilnya. Hasil yang diperlihatkan Rapid Miner pun dapat ditampilkan secara visual dengan grafik. Menjadikan RapidMiner adalah salah satu software pilihan untuk melakukan ekstraksi data dengan metode-metode data mining Brilian Rahmat C.T.I. (I et al., 2017).

METODOLOGI PENELITIAN

Penelitian menggunakan algoritma *Naive Bayes* dan dalam menghitung performa dari algoritmanya digunakan *software Rapid Miner*. Data yang digunakan adalah data set penyakit hepatitis dan termasuk data sekunder karena diperoleh dari UCI (Universitas California, Invene) *Machine Learning Repository* yang ada pada web https://archive.ics.uci.edu/ml/datasets/Hepatitis .

Berikut langkah-langkah yang digunakan dalam penelitian ini:



Gambar 2. Langkah-Langkah Penelitian

1. Pengumpulan Data
Teknik pengumpulan data merupakan langkah yang paling strategis dalam penelitian, karena tujuan utama dari penelitian adalah mendapatkan data (Sugiyono, n.d.). Dan terdapat jenis sumber dalam pengumpulan data, yaitu Data Primer yang artinya data yang diperoleh secara langsung dari objek penelitian, sedangkan Data Sekunder adalah data yang diperoleh dari *literature*, buku referensi, maupun browsing internet (Andi Prastowo, 2012). Dan sumber data yang digunakan dalam penelitian ini disebut data sekunder, karena data ini diperoleh dari UCI *Repository* dengan alamat web https://archive.ics.uci.edu/ml/datasets/Hepatitis .

Data tersebut didonasikan oleh G.Gong (Carnegie-Mellon University) via Bojan Cestnik di Yugoslavia pada 01 November 1988. Jumlah data yang ada sebanyak 155 data dengan 123 pasien penyakit hepatitis yang *Life* dan 32 pasien penyakit hepatitis yang *Die*. Atribut yang terdapat pada dataset hepatitis ini yaitu: *Age, Sex, Steroid, Antivirals, Fatigue, Malaise, ANorexia, Liver Big, Liver Firm, Spleen Palpable, Spiders, Ascites, Varices, Bilirubin, Alk Phosphate, Sgot, Albumin, Protime, Histology*, dan *Class* (atribut hasil prediksi).

2. Pengolahan Data
Teknik yang digunakan dalam pengolahan data ini adalah menggolongkan data berdasarkan hasil dari atribut *Class*. Dan data yang diperoleh adalah sebanyak 155 data dengan 123 berada di *Class* "*Life*" dan 32 data berada di *Class* "*Die*".

3. Metode Yang Digunakan
Metode yang digunakan dalam penelitian ini adalah metode klasifikasi data mining algoritma *Naive Bayes*. Perhitungan manual menggunakan *Excel* dan juga pengujian model menggunakan Aplikasi *Rapid Miner* dan nantinya akan diakurasikan apakah hasilnya akan sama atau berbeda dalam memprediksi penyakit hepatitis.

HASIL DAN PEMBAHASAN

Pada tahap ini dilakukan pengujian metode *Naive Bayes* yang akan digunakan untuk memprediksi penyakit hepatitis. Langkah-langkah yang dilakukan adalah untuk menghitung nilai probabilitas nilai "*Life*" dan "*Die*" dari masing-masing atribut pada total kasus "*Life*" dan "*Die*" dari keseluruhan data.

Tabel 1. Perhitungan Probabilitas Prior Keseluruhan

Atribut	Jumlah	*Life*	*Die*	P(X|Ci)	
				Life	*Die*
Total Kasus	155	123	32	0,794	0,206
Age 20-40	82	73	9	0,890	0,110
Age 41-60	59	39	20	0,661	0,339
Age 61-80	14	11	3	0,786	0,214
Steroid = No	76	56	20	0,737	0,263
Steroid = Yes	79	67	12	0,848	0,152
Malaise = No	61	38	23	0,623	0,377
Malaise = Yes	94	85	9	0,904	0,096
Liver Big = No	25	22	3	0,880	0,120
LiverBig= Yes	130	101	29	0,777	0,223
Spiders = No	51	29	22	0,569	0,431
Spiders = Yes	104	94	10	0,904	0,096
Varices = No	18	7	11	0,389	0,611
Varices = Yes	137	116	21	0,847	0,153

Untuk menentukan kelas manakah yang akan digunakan dalam perhitungan selanjutnya, maka data yang akan digunakan hanyalah sample dari data yang ada. Data tersebut disebut data probabilitas posterior karena datanya bersumber dari data prior yang telah dihitung sebelumnya.

Tabel 2. Probabilitas *Posterior*

Atribut	Nilai	*Life*	*Die*
Total Kasus	155	0,794	0,206
Age	20-40	0,890	0,110
Steroid	*No*	0,737	0,263
Malaise	*No*	0,623	0,377
Liver Big	*Yes*	0,777	0,223
Spiders	*Yes*	0,904	0,096
Varices	*No*	0,389	0,611

Berdasarkan tabel di atas, maka dapat dihitung probabilitas dari tiap atribut yang ada.

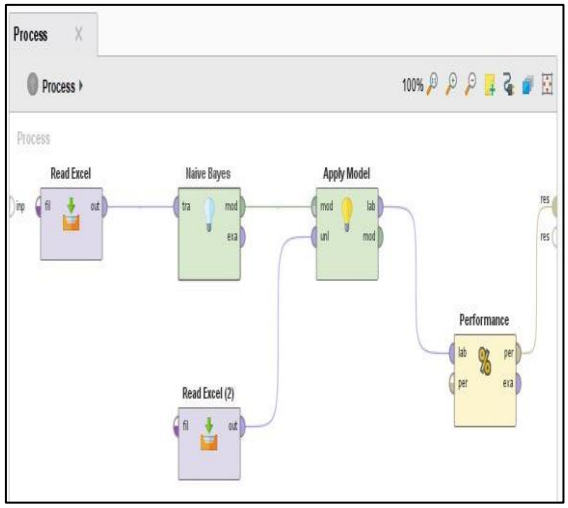
1. Perhitungan Probabilitas "*Life*" untuk setiap atribut:
   = 0,794*0,890*0,737*0,623*0,777*0,904*0,389
   = 0,0885541

2. Perhitungan Probabilitas "*Die*" untuk setiap atribut:
= 0,206*0,110*0,263*0,377*0,223*0,096*0,611
= 2,9472E-05

3. Perbandingan Probabilitas Antara "*Life*" dan "*Die*"
Probabilitas *Life* = 0,0885541
Probabilitas *Die* = 2,9472E-05

Dikarenakan 0,0885541 lebih besar dari 2,9472E-05, maka dapat disimpulkan bahwa data testing tersebut termasuk klasifikasi "*Life*".

Pengujian Probabilitas Menggunakan *Rapid Miner*:



Gambar 3. Proses Pengujian Data Menggunakan *Naive Bayes*

Berikut hasil perhitungan akurasi data training menggunakan *Naive Bayes*. Dapat diketahui tingkat akurasinya sebesar 76.77%. Dari 155 data sebanyak 88 data diprediksikan sesuai yaitu 88 data "*Life*" dan 1 data yang diprediksikan "*Life*" tapi ternyata hasilnya "*Die*". Dan sebanyak 35 data diprediksi "*Die*" ternyata termasuk klasifikasi "*Life*" dan sebanyak 31 data diprediksi sesuai yaitu "*Die*".

Tabel 3. Tabel *Accuracy Naive Bayes*

accuracy: 76.77%

	true Life	true Die	class precision
pred. Life	88	1	98.88%
pred. Die	35	31	46.97%
class recall	71.54%	96.88%	

KESIMPULAN

1. Metode *Naive Bayes* termasuk algoritma yang akurat untuk memprediksi karena hasil akurasi menggunakan *Rapid Miner* menunjukkan lebih dari 50% yaitu sebesar 76.77%. Dengan hasil *Class Precision* yang tertinggi sebesar 98.88% untuk prediksi "*Life*", dan *Class Recall* sebesar 96.88% untuk Prediksi "*Die*".
2. Dapat digunakan metode algoritma klasifikasi yang lain untuk mengembangkan penelitian ini seperti metode C4.5, KNN, SVM, dan lain-lain.
3. Dapat juga penelitian ini menggunakan dua atau lebih metode algoritma sekaligus untuk mengetahui algoritma manakah yang hasilnya lebih akurat.

REFERENSI

Andi Prastowo. (2012). *Metode Penelitian Kualitatif dalam Perspektif Rancangan Penelitian*. Yogyakarta.

Bari, M., Sitorus, S. H., & Ristian, U. (2018). IMPLEMENTASI METODE NAÏVE BAYES PADA APLIKASI PREDIKSI PENYEBARAN WABAH PENYAKIT ISPA (Studi Kasus : Wilayah Kota Pontianak), *06*(03), 205–214.

Bustami. (2014). PENERAPAN ALGORITMA NAIVE BAYES, *8*(1), 884–898.

Gorunescu, F. (2011). *Data Mining: Concepts and Techniques. Verlag berlin Heidelberg: Springer*.

I, B. R. C. T., Gafar, A. A., Fajriani, N., Ramdani, U., Uyun, F. R., P, Y. P., & Ransi, N. (2017). Implemetasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan. *Seminar Nasional Riset Kuantitatif Terapan*, (April), 58–62.

Jananto, A. (2013). Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa. *Teknologi Informasi*, *18*(1), 9–16.

Kajian Pustaka. (n.d.). Bab ii kajian pustaka, 13–36.

Kementrian Kesehatan RI : 2014. (n.d.). Daftar Pustaka 1 infodatin-hepatitis.pdf. https://doi.org/24427659

Larose, D. T. (2005). *Discovering Knowledge in Databases. New Jersey: John Willey & Sons Inc.*

Muslehatin, W., & Ibnu, M. (2017). Penerapan Naïve Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau, 18–19.

Prasetyo, E. (2012). *Data Mining: Konsep Dan Aplikasi Mengunakan Matlab*.

Prayoga, N. D., Hidayat, N., & Dewi, R. K. (2018). Sistem Diagnosis Penyakit Hati Menggunakan Metode Naïve Bayes, *2*(8), 2666–2671.

Sabransyah, M., Nasution, Y. N., & Tisna, D. (2017). Aplikasi Metode Naive Bayes dalam

Prediksi Risiko Penyakit Jantung Naive Bayes Method for a Heart Risk Disease Prediction Application. *Jurnal EKSPONENSIAL*, *8*, 111–118.

Sugiyono. (n.d.). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*.

Suryanih. (2011). Diagnosis Kesulitan Belajar Matematika Siswa Dan Solusinya Dengan Pembelajaran Remedial 2011 M / 1432 H.

PROFIL PENULIS

Deny Novianti, S.Kom
Lahir di Bekasi, 05 November 1994. Adalah salah satu Mahasiswa Berprestasi BSI yang kini bekerja di Bagian Pengembangan Dosen BSI sejak tahun 2017 – sekarang. Saat ini penulis sedang melanjutkan Studi S2 di STMIK Nusa Mandiri Jakarta jurusan Ilmu Komputer.