

Optimasi Metode *Naïve Bayes* Menggunakan Algoritma *Particle Swarm Optimization* (PSO) Untuk Prediksi Penyakit Diabetes Mellitus

Moh. Sukron¹, Ahmad Supriadi², Rizal Sulton³
^{1,2,3} Universitas Nurul Jadid, Indonesia

Info Artikel

Riwayat Artikel

Diterima: 15-11-2021

Disetujui: 17-12-2021

Kata Kunci

Diabetes Mellitus

Naive Bayes

Data Mining

Particle Swarm Optimization
(PSO)

e-mail*

moh.sukron12012752@gmail.com

ABSTRAK

Diabetes Mellitus merupakan salah satu penyakit dimana terdapat tingkat kadar gula (glukosa) yang tinggi didalam darah dalam konteks defisiensi insulin relatif dan resistensi insulin. Diabetes Mellitus sulit untuk diprediksikan. Dengan pemanfaatan database pasien, peneliti mencoba mengolah variabel pasien Diabetes Mellitus dalam mendapatkan akurasi prediksi penyakit Diabetes Mellitus dengan memanfaatkan data mining. Metode yang digunakan adalah Naive Bayes, namun metode Naive Bayes memiliki kelemahan yaitu hasil probabilitas kurang berjalan seara optimal dan sering salah pada atribut. Guna mengatasi kelemahan Naive Bayes, salah satu cara yaitu dengan metode pembobotan atribut menggunakan algoritma Particle Swarm Optimization (PSO). Hasil penelitian menunjukkan akurasi metode naive bayes menghasilkan akurasi sebesar 86,80% dan Naive Bayes dengan Particle Swarm Optimization (PSO) menghasilkan akurasi 89,84%. Dari hasil tersebut dapat disimpulkan metode naive bayes yang di optimasi dengan pembobotan atribut di dapatkan nilai akurasi yang lebih baik dengan peningkatan akurasi 3,04%.

1. PENDAHULUAN

Diabetes Mellitus merupakan salah satu penyakit dimana terdapat tingkat kadar gula (glukosa) yang tinggi didalam darah dalam konteks defisiensi insulin relatif dan resistensi insulin. Terdapat tiga jenis Diabetes Mellitus, Diabetes Mellitus tipe 1, Diabetes Mellitus tipe 2 dan Diabetes gestasional. Diabetes Mellitus tipe 1 dikenal dengan istilah insulin dependent diabetes. Istilah tersebut mengacu pada keterkaitan penyakit diabetes terhadap ketidakmampuan tubuh untuk memproduksi insulin. Sedangkan Diabetes Mellitus tipe 2 sering juga disebut sebagai penyakit kencing manis. Pada penderita Diabetes Mellitus tipe 2, organ pankreas memproduksi hormon insulin yang terbatas sehingga kadar gula dalam darah meningkat tinggi (hiperglikemia). Pada umumnya gejala klasik Diabetes Mellitus tipe 2 ditandai dengan haus berlebihan (polidipsi), sering mengeluarkan urin (poliurin) dan rasa lapar terus menerus (polifagi). Dan yang terakhir adalah Diabetes Mellitus Gestasional, dimana jenis penyakit ini erat kaitannya pada wanita pada saat kondisi kehamilan. Melonjaknya kadar gula darah ibu hamil menjadi penyebab utama terjadinya penyakit ini[1].

Diabetes Mellitus termasuk penyakit kronis yang berkembang secara bertahap, sehingga dapat memicu sejumlah komplikasi penyakit lainnya jika tidak segera ditangani dengan baik. Sejumlah komplikasi yang dialami oleh penderita Diabetes Mellitus antara lain penyakit jantung, stroke, aterosklerosis, gagal ginjal, retinopati yang mempengaruhi

penglihatan mata dan kurangnya sirkulasi dibagian tungkai yang mengharuskan dilakukannya amputasi[2].

Beberapa hal yang meningkatkan faktor resiko deabetes millitus antara lain obesitas, umur, makanan, genetik, hipertensi, jenis kelamin, merokok, konsumsi alkohol, serta kurangnya aktifitas tubuh. Menurut data International Diabetes Federation (IDF) tahun 2013, sebanyak 382 juta orang di dunia mengalami Diabetes. Sedangkan di Indonesia berdasarkan data Riskendas tahun 2013, jumlah penduduk yang mengalami diabetes dengan usia 15 tahun keatas sebanyak 12.191.564 orang [3]. Pengetahuan yang minim tentang penyakit Diabetes Millitus baik dari segi penyebab serta tanda-tanda terjadinya penyakit ini menjadi faktor resiko utama [4].

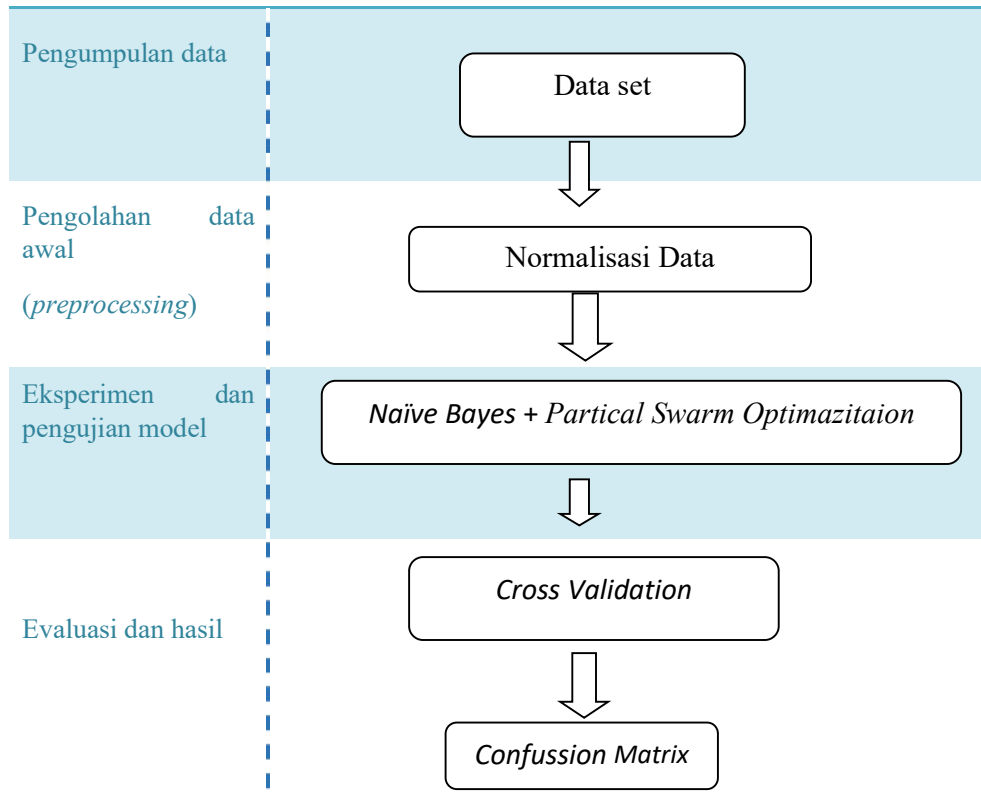
Era global ini penfaatan big data dalam berbagai bidang telah banyak dilakukan untuk mendapat informasi yang bermanfaat. Salah satunya dengan cara data mining. Seperti yang telah di ketahui, bahwa data mining merupakan sebuah proses yang menggunakan tehnik machine learning, Artificial Intelligence serta model statistik matematika guna meng-ekstraksi dan menggali sebuah informasi yang bermanfaat dari database yang besar (big data)[5]. Pada dasarnya data mining memiliki 4 fungsi dasar diantaranya fungsi prediksi, fungsi deskripsi, fungsi klasifikasi dan fungsi asosiasi. Klasifikasi data minning merupakan sebuah proses yang berfungsi untuk mencari sebuah model yang menggambarkan sebuah konsep data. Proses tersebut berguna untuk menjelaskan data serta berfungsi untuk meramalkan kecenderungan data di masa depan. Dan fungsi klasifikasi dengan pemanfaatan algorima atau metode data mining alternatif menjadi solusi yang tepat dalam memprediksi berbagai penyakit, salah satunya kegiatan dalam memprediksikan penyakit Diabetes Millitus. Pemanfaatan metode atau algoritma menjadi titik utama dalam mendapatkan nilai prediksi yang baik.

Menurut penelitian [6], tujuan menerapkan metode Naïve Bayes untuk memperediksi kemungkinan komplikasi penyakit diabetes terhadap penyakit jantung. Algoritma yang digunakan dalam penelitian ini menggunakan Naïve Bayes dengan hasil prediski yang optimal menggunakan minimum training set. Dataset yang digunakan berasal dari data set klinik yang di kumpulkan dari diabetec research institute di Chennai. Dari hasil penelitian didapatkan bahwasanya atribut pasien dapat memperediksi komplikasi penyakit diabetes terhadap penyakit jantung walaupun atribut tersebut bukan merupakan indikator langsung penyakit jantung. Dan dari hasila lgoritma di dapatkan nilai akursi prediksi Naïve Bayes 71,5% dan Bayes Network 72,3%.

Naïve Bayes adalah salah satu metode klasifikasi yang efektif dan mudah serta banyak diuji untuk induksi probabilitas yang dikenal dengan nama lain bayesian classifier [7]. Meskipun paling banyak di gunakan tapi Naïve Bayes masih memiliki kelemahan yaitu hasil probabilitas kurang berjalan seara optimal dan sering salah pada atribut. Guna mengatasi kelemahan Naïve Bayes, salah satu cara yaitu dengan metode pembobotan atribut agar akurasi Naïve Bayes meningkat[8]. Salah satu algoritma dalam menyelesaikan permasalahan tersebut adalah dengan pemanfaatan algoritma Particle Swarm Optimization (PSO). Algoritma Particle Swarm Optimization (PSO) dapat digunakan untuk meningkatkan hasil akurasi metode Naïve Bayes dengan cara pembobotan atribut. Data yang digunakan adalah data yang bersumber dari UCI repository dataset, berupa data Z-Alizadeh sani dataset sebanyak 303 record.

Dari pemaparan diatas, penelitin ini dilakukan untuk memperediksi penyakit Diabetes Mellitus. Penggunaan metode Naïve Bayes serta optimasi Particle Swarm Optimization (PSO) diharapkan dapat menghasilkan nilai akurasi yang tinggi. Pun dengan akurasi yang tinggi menjadikan hasil prediksi semakin tepat sehingga dapat menghasilkan output yang bagus pula. Penelitian ini diharapkan mampu memberikan manfaat kepada pihak yang menangani penderita Diabetes Millitus. Sehingga, dengan adanya prediksi tersebut pihak terkait dapat memudahkan dalam mengambil keputusan yang tepat dalam menangani penderita Diabetes Millitus.

2. METODE



Data set yang digunakan dalam penelitian ini bersumber dari UCI repository dataset, berupa data Z-Alizadeh sani dataset sebanyak 303 record. variabel yang digunakan pada data penyakit deabetes millitus sebanyak 8 variabel dengan rincian 7 atribut dan 1 label, di jenis atribut sbegaimana pada tabel berikut :

No	Atribut	Jenis atribut	Keterangan
1	Age	Numeric	-
2	Sex	Nominal	-
3	BMI	Numeric	Body Mass Index
4	Hypertension	Nominal	-
5	FBS	Numeric	Fasting Blood Sugar (mg/dL)
6	Blood Preasure	Numeric	Hemoglobin
7	HB	Numeric	-
8	DM	nominal	Label/ Deabetes Millitus

2.1. Pengolahan data awal

Untuk mendapatkan data yang berkualitas digunakan tahap tahap persiapan data (data preparation) karena tidak semua data langsung dapat digunakan. Maka dalam penelitian ini dilakukan preprocessing diantaranya:

- a. Trasnformasi discretization: metode yang digunakan adalah unsupervised discretization dengan pendekatan equal width. Dimana setiap atribut numeric akan di bagi sesuai interval data yang telah di tentukan pada diatas.

- b. Transformasi binarization:
Beberapa atribut dilakukan transformasi menjadi $[0,1]$. Atribut yang di ubah terdapat pada atribut nominal seperti pada Diabetes Mellitus, jenis kelamin, merokok, dan olah raga
- c. Normalisasi Standar Devisiasi
Agar jangkauan nilai setiap atribut tidak terpaut jauh maka dilakukan normalisasi data dengan metode Standar Devisiasi dengan menerapkan formula 2.2. Atribut yang dinormalisasi berupa atribut numerik. Diantaranya atribut Umur, Berat Badan, Tinggi Badan, Kadar Gula Puasa – 2 jam

Langkah selanjutnya dataset yang telah melalui proses preprocessing tersebut akan dilakukan eksperimen (Processing) dengan menerapkan algoritma Naïve Bayes dan pengujian model menggunakan 10 fold cross-validation.

2.2. Eksperimen dan Pengujian Model.

Pada penelitian ini digunakan 2 metode yaitu metode Naïve Bayes dan metode Naïve Bayes yang dikombinasikan dengan optimasi Partical Swarm Optimization (PSO) sebagai pembobotan atribut. Sebelum dataset di terapkan dalam model Naïve Bayes dan Partical Swarm Optimization (PSO), data terlebih dahulu di validasi dengan 10fold cross validation serta dengan hipotesa hasil berupa akurasi yang tinggi. Pemaparan detailnya sebagai berikut :

A. Naïve Bayes

Algoritma yang digunakan dalam penelitian ini adalah Naïve Bayes. Pada metode naïve byes semua atribut akan memberikan kontribusinya dalam mengambil keputusan dengan bobot yang sama penting dan setiap atribut saling bebas satu sama lain. Berdasarkan dataset , akan dibangun sebuah keputusan untuk menentukan label diagnosi penyakit diabetes mellitus. Label akan di bagi menjadi 2 hasil tes, nilai “y” tested_positive (terdiagnosis) dan nilai “t” untuk tested_negative (non terdiagnosis). Algoritma Naïve Bayes dapat dijelaskan dalam sekerna berikut:

- a) Membaca data latih Diabetes Mellitus yang telah melalui preprocessing .
- b) Menghitung jumlah probabilitas setiap atribut terhadap output/label (Diabetes Mellitus).
- c) Mengalikan setiap probabilitas atribut terhadap label Diabetes Mellitus nilai probabilitas “y” tested_positive (terdiagnosis) dan nilai “y” tested_positive (terdiagnosis) dan nilai probabilitas nilai “t” untuk tested_negative (non terdiagnosis).
- d) Membandingkan dan mengambil nilai terkecil dari hasil probabilitas label Diabetes Mellitus nilai probabilitas “y” tested_positive (terdiagnosis) dan nilai “y” tested_positive (terdiagnosis) dan nilai probabilitas nilai “t” untuk tested_negative (non terdiagnosis) sebagai hasil akhir prediksi.

B. Naïve Bayes dalam Partical Swarm Optimization (PSO)

Untuk meningkatkan kinerja Naïve Bayes dalam penelitian ini akan dioptimasi dengan parcal swarm optimization . Algoritma Naïve Bayes dan Partical Swarm Optimization (PSO) akan diimplementasikan sebagai berikut [17]:

- a) Menginisialisai setiap atribut
- b) Membandingkan kinerja masing masing individu dengan kinerja terbaik
- c) Membandingkan kinerja setiap partikel denganglobal best partikel.
- d) Jika perhitungan bobot telah selesai maka hitung bobot yang lainnya dengan kembali ke langkah 1 hingga selesai.
- e) Menghitung rata-rata bobot.

Setelah mendapatkan nilai bobot setiap atribut, maka penggunaan atribut yang mendapatkan pembobotan kecil dapat diseleksi atau diabaikan sehingga dapat mengoptimalkan hasil yang diinginkan.

C. Evaluasi dan Hasil.

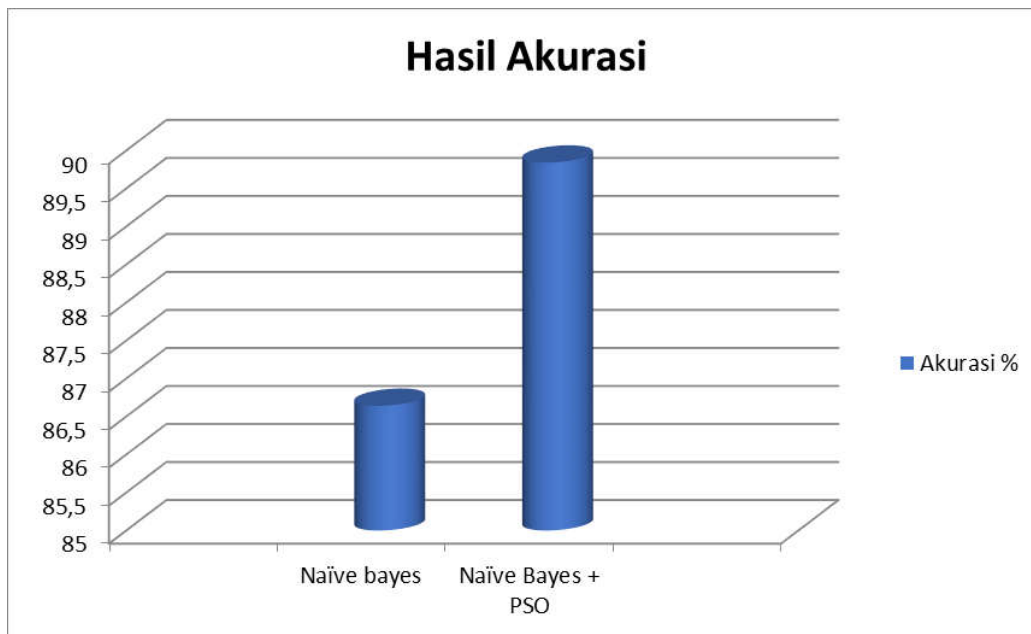
Cross validation merupakan sebuah tindakan pembuktian dari sebuah metode atau performa suatu algoritma. Pembagian paling banyak digunakan dalam penelitian klasifikasi data mining merupakan dengan membagi data secara acak menjadi 10 bagian. Satu bagian menjadi data testing dan 9 bagian dijadikan data training. Validasi yang seperti ini disebut juga dengan 10fold cross validation [18].

Dalam penelitian ini dilakukan pengujian dipilih 10-fold cross validation yakni dengan membagi rata data secara acak menjadi 10 bagian kemudian. Data yang digunakan sebanyak 303 record sehingga sekitar 30 data akan dijadikan data testing dan 273 sisanya dijadikan data training. Proses seperti itu akan terus berlangsung satu persatu setiap bagian dari hingga 9 bagian lainnya bergantian menjadi data testing. Tahap akhirnya merupakan hasil rata-rata akurasi dari 10 kali percobaan tersebut.

Dalam Pengukuran kinerja algoritma Naive Bayes untuk metode klasifikasi menggunakan Confusion Matrix. Confusion Matrix adalah tabel matriks yang terdiri dari dua kelas, yaitu kelas yang di anggap positif dan kelas yang dianggap sebagai negatif. Confusion matrix memiliki tabel yang berisi tentang informasi data aktual dan data prediksi pada sebuah kasus data yang diklasifikasi. Dalam confusion matrix, total record yang dipakai dalam dataset baik yang diprediksikan kedalam kelas positif Diabetes Mellitus (y) ataupun kelas negatif Diabetes Mellitus (t).

3. HASIL DAN PEMBAHASAN

Dari hasil pengujian metode *Naive Bayes* dan metode *Naive Bayes* yang di optimasi dengan pembobotan atribut didapatkan nilai akurasi yang lebih baik dengan peningkatan akurasi sekitar 3,04%. Dengan rincian metode *naive bayes* menghasilkan akurasi sebesar 86,80% dan *Naive Bayes* dengan *Particle Swarm Optimization* (PSO) menghasilkan akurasi 89,84. Hasil dari kedua metode tersebut di tampilkan dalam histogram berikut



Gambar 1. Berikan judul gambar yang jelas dan deskriptif

4. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan, metode Naive Bayes dengan pembobotan atribut metode Particle Swarm Optimization (PSO) mendapatkan nilai akurasi yang lebih baik dari pada menggunakan metode Naive Bayes saja. Metode Naive Bayes menghasilkan akurasi

sebesar 86,80% dan Naive Bayes dengan Particle Swarm Optimization (PSO) menghasilkan akurasi 89,84%. Dari hasil tersebut dapat disimpulkan metode Naive Bayes yang di optimasi dengan pembobotan atribut Particle Swarm Optimization (PSO) di dapatkan nilai akurasi yang lebih baik dengan peningkatan akurasi sekitar 3,04%.

Proses penelitian ini memiliki beberapa kendala baik dari segi dataset maupun perangkat keras yang digunakan. Untuk penelitin selanjutnya terdapat beberapa saran sebagai berikut:

- a) Penelitian ini mengkomparasikan algoritma Naive Bayes dan Naive Bayes yang dikombinasikan dengan algoritma Particle Swarm Optimization (PSO) untuk pembobotan atribut. Untuk penelitian selanjutnya dapat dikembangkan dengan menggunakan algoritma yang lainnya seperti C4.5, SVM ataupun Neural Network. Untuk optimasi bisa digunakan Genetic Algorithm ataupun Ant Colony Optimization.
- b) Penelitian ini menggunakan UCI repository dataset, berupa data Z-Alizadeh sani dataset sehingga untuk penelitin selanjutnya mungkin bisa dibandingkan dengan dataset penyakit Diabetes Mellitus dari sumber lainnya.
- c) Semakin banyak atribut yang digunakan semakin bagus pula hasil penelitian sehingga untuk penelitian selanjutnya mungkin dapat ditambahkan beberapa atribut lagi sehingga bisa dilakukan seleksi atribut terlebih dahulu sebelum dilakukan pembobotan atribut untuk prediksi penyakit Diabetes Mellitus ataupun penyakit lainnya.

5. DAFTAR PUSTAKA

- [1] E. Dwi Amelisa Edwina, Asman Manaf, “Artikel Penelitian Pola Komplikasi Kronis Penderita Diabetes Melitus Tipe 2 Rawat Inap di Bagian Penyakit Dalam RS . Dr . M . Djamil Padang Januari 2011 - Desember 2012,” vol. 4, no. 1, 2015.
- [2] G. Allen, “Managing acute pulmonary oedema,” vol. 40, no. 2, pp. 59–63, 2017.
- [3] Depkes, “Infodatin-Situasi dan Analisis Diabetes.” Jakarta Selatan, 2014.
- [4] J. Inovasi, “Pencegahan Penyakit Diabetes Mellitus Melalui Program Penyuluhan Dan Pemeriksaan Kadar Gula,” vol. 3, no. 3, pp. 180–185, 2014.
- [5] J. P. Jiawei han, Micaheline Kamber, *Data Mining: Concepts and Techniques, 3rd Edition*.
- [6] G. Parthiban and C. A. H. College, “Diagnosis of Heart Disease for Diabetic Patients using Naive Bayes Method,” vol. 24, no. 3, pp. 7–11, 2011.
- [7] B. Classification, “Naive-Bayes Classification Algorithm.”
- [8] N. A. Zaidi, M. J. Carman, and G. I. Webb, “Alleviating Naive Bayes Attribute Independence Assumption by Attribute Weighting,” vol. 14, pp. 1947–1988, 2013.
- [9] H. Muhamad *et al.*, “Optimasi Naive Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris,” vol. 4, no. 3, pp. 180–184, 2017.
- [10] R. M. Rahman and F. Afroz, “Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis,” vol. 2013, no. March, pp. 85–97, 2013.
- [11] V. Karthikeyani, “Comparison a Performance of Data Mining Algorithms (CPDMA) in Prediction Of Diabetes Disease,” vol. 5, no. 3, pp. 205–210, 2013.
- [12] P. K. D. bS. Mythili Thirugnanama, “Improving the Prediction Rate of Diabetes Diagnosis Using Fuzzy, Neural Network, Case Based (FNC) Approach,” vol. 38, pp. 1709–1718, 2012.
- [13] D. H. C. Pangemanan, “Hubungan diabetes melitus dengan kualitas tidur,” vol. 4,

2016.

- [14] Florin Gorunescu, *Data Mining : Concepts, Models and technigues*. 2010.
- [15] A. Kaveh, *Advances in Metaheuristic Algorithms for Optimal Design of Structures-Chapter 2 (PSO)*. 2017.
- [16] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
- [17] J. Li, L. Ding, and B. Li, “A Novel Naive Bayes Classification Algorithm Based on Particle Swarm A Novel Naive Bayes Classification Algorithm Based on Particle Swarm Optimization,” no. December, 2014.
- [18] ian H witten, *Data Mining Practical Machine Learning Tools and Techniques*.