

PENERAPAN TEKNIK BAGGING UNTUK MENINGKATKAN AKURASI KLASIFIKASI PADA ALGORITMA NAIVE BAYES DALAM MENENTUKAN BLOGGER PROFESIONAL

Andi Kurniawan, M.Kom.

Sistem Informasi, Universitas Muhammadiyah Kudus
Jln. Ganesha I Purwosari Kudus 59316, Jawa Tengah, Indonesia
Kurniawan_kh@yahoo.com

Abstrak

Blog memang banyak diminati baik dari kalangan remaja hingga dewasa maupun orang tua. Blogger memiliki latar belakang yang sama yaitu suka menulis dengan memanfaatkan internet. Satu hal yang membuat blog akan disukai pengunjung yaitu blog yang konsisten. Sebagai blogger yang bertanggung jawab, menulis atau mengisi blog haruslah konsisten. Konsisten dalam pengertian waktu dan tema. Konsisten waktu artinya blogger harus menulis rutin misalnya setiap hari, seminggu sekali atau dua kali, sebulan sekali, dan lainnya. Sebenarnya bebas tetapi jika rutin atau terjadwal maka akan jadi blogger profesional. Data mining dapat dimanfaatkan untuk mengklasifikasikan blogger profesional. Ada beberapa model data mining salah satunya klasifikasi. Di bidang klasifikasi, ada banyak cabang yang berkembang yaitu pohon keputusan (decision tree). Metode yang digunakan untuk mengolah data yang sifatnya besar untuk menemukan pola yang terdapat didalamnya diantaranya adalah Algoritma Naïve Bayes Classification (NBC) yang merupakan algoritma dari metode teorema Bayes. Dalam riset ini, data yang digunakan adalah Blogger dataset yang diambil dari UCI repository of machine learning. Pada dataset ini atributnya terdiri: Pendidikan (degree), tingkah politik (caprice), topik, media local turnover (LMT) dan ruang lokal, politik dan sosial (LPSS). Akurasi sangat penting dalam pengklasifikasian. Ensemble method adalah metode yang digunakan untuk meningkatkan akurasi algoritma klasifikasi dengan membangun beberapa classifier dari data training. Dari hasil penelitian, dengan menerapkan teknik bagging untuk klasifikasi berbasis ensemble pada algoritma naïve Bayes dapat meningkatkan akurasi sebesar 3,00%. Dengan akurasi awal 77,00%, setelah diterapkan teknik bagging menjadi 80,00%.

Kata Kunci: Data mining, Naïve Bayes Classification, NBC, Ensemble, Bagging, blogger profesional.

Abstract

Blogs are in great demand both from adolescents to adults and parents. Bloggers have the same background, they like to write using the internet. One thing that makes a blog will be liked by visitors is a consistent blog. As a responsible blogger, writing or filling a blog must be consistent. Be consistent in terms of time and theme. Consistent time means that bloggers must write routinely for example every day, once a week or twice, once a month, and others. Actually free but if it's routine or scheduled it will become a professional blogger. Data mining can be used to classify professional bloggers. There are several data mining models, one of which is classification. In the field of classification, there are many branches that develop, namely the decision tree. The method used to process large data to find patterns contained therein is the Naïve Bayes Classification (NBC) algorithm which is an algorithm of the Bayes theorem method. In this research, the data used is the Blogger dataset taken from the UCI repository of machine learning. In this dataset the attributes consist of: education (degree), political behavior (caprice), topics, local media turnover (LMT) and local, political and social space (LPSS). Accuracy is very important in classification. Ensemble method is a method used to improve the accuracy of classification algorithms by building several classifiers from training data. From the results of the study, applying bagging techniques for ensemble-based classification on the naïve Bayes algorithm can increase accuracy by 3.00%. With an initial accuracy of 77.00%, after applying the bagging technique to 80.00%.

Keywords: Data mining, Naïve Bayes Classification, NBC, Ensemble, Bagging, professional bloggers.

I. PENDAHULUAN

Blog bisa disebut buku harian di internet yang lahir sebagai situs web dinamis dan melihat asalnya pada tahun 1997 berkat **Dave Winer**. Dave Winer adalah pengembang perangkat lunak pertama yang menerbitkan artikel baru secara mandiri. Blog dari segi fungsinya di bagi menjadi blog pribadi dan blog bisnis

Blog pribadi adalah Anda bebas untuk mengungkapkan pikiran Anda untuk membiarkan kata-kata mengalir di selembar kertas. Blogging pribadi adalah cara spektakuler untuk menyuarakan kreativitas Dan untuk membuat orang membaca karya Anda. Dengan cara ini Anda dapat mengekspos bakat, Anda dapat memastikan bahwa **ada umpan balik** dari pembaca .

Setiap bentuk kreativitas membutuhkan perbandingan. Blogger jika ingin tahu apakah yang ditulis itu berharga dan berhasil di baca para pembaca , itu cara melihat bahwa blogger kita berhasil atau tidak.

Blog perusahaan **adalah investasi** . Dalam hal ini banyak menyebut tentang aktivitas freelancer, freelance, yang ingin menggunakan blog untuk mempromosikan bisnisnya . Uang yang di dikeluarkan bukan biaya tapi cara untuk menghasilkan uang lebih banyak dalam jangka panjang. Dalam sebuah bloger dapat terlihat blog yang banyak disukai oleh pengunjung biasanya pembuat atau blogernya sudah profesional. Blogger yang profesional selalu konsisten. Kekonsistennya dapat berupa pengisian konten dan waktu. Pengunjung akan selalu terikat akan konten yang diminatinya dan tergantung pada keprofesionalan bloggernya.

Di era teknologi Komputer yang terus berkembang dengan pesat dapat membantu kita untuk mendeteksi blogger profesional secara akurat dan dapat menghemat waktu untuk tujuan sebuah bisnis . Data mining adalah bidang dalam ilmu komputer digunakan untuk prediksi diberbagai bidang kehidupan dan berbagai tujuan. Semua ini adalah proses menemukan dataset baru dari dataset yang sebelumnya diketahui melalui analisis dataset untuk sebuah tujuan (Larose, 2006). Untuk memprediksi blogger profesional dan tidak dengan melalui data mining diperlukan unsur – unsur penunjang

dalam penentuannya disertai dengan data yang valid.

Salah satu algoritma *data mining* adalah Naïve Bayes. Metode Naïve Bayes digunakan mengklasifikasikan penentuan keprofesionalan seorang blogger.

Penelitian dilakukan untuk menganalisis melalui algoritma Naïve Bayes untuk mengklasifikasi penentuan keprofesionalan seorang blogger sehingga mendapatkan akurasi yang di dapat dari hasil proses evaluasi.

II. LANDASAN TEORI

A. Data Mining

Data mining adalah sebuah ilmu yang mempelajari metode untuk menggali data yang besar untuk menemukan sebuah pola data yang bernilai pengetahuan sehingga data tersebut dapat digunakan untuk sebuah tujuan. Data Mining dapat disimpulkan mempunyai 3 inti penting (Larose, 2006) yaitu :

Data adalah fakta yang terekam dan belum mempunyai arti.

Informasi merupakan ekstraksi data yang sudah diolah

Pengetahuan dimana pengetahuan berupa pola dan aturan atau model yang muncul dari ekstraksi data

Nama lain dari data mining adalah Knowledge Discovery in Data, ada juga tentang, Knowledge Extraction, Big data atau bussines intelijen dan pattern analisis atau information harvesting .

Konsep dari data mining yaitu himpunan data yang telah terhimpun banyak sekali kemudian dengan metode data mining yang ada diproses dengan rumus/algoritma data-data tersebut diambil atau diekstraksi untuk menjadi di sebuah pengetahuan. Pengetahuan Itulah bisa untuk diambil untuk sebuah tujuan tertentu. Pengaturan tersebut sangat berguna untuk berbagai keperluan .

B. Algoritma Naïve Bayes

Algoritma Naïve Bayes akan mengevaluasi setiap atribut yang berkontribusi prediksi pada atribut target. Naïve Bayes tidak memperhitungkan relasi antar atribut-atribut kontributor prediksi, tidak seperti Decision Tree yang memperhitungkan relasi antara atribut. Bentuk tugas dasar yang dilakukan oleh algoritma Naïve Bayes adalah hanyalah

klasifikasi (Boukenze *et al.*, 2012) . Naïve Bayes merupakan teknik data mining dengan pendekatan teori probabilitas untuk membangun sebuah model klasifikasi berdasarkan pada kejadian masa lalu yang mempunyai potensi membentuk sebuah objek baru yang dikategorikan sebagai kelas yang memiliki probabilitas terbaik.

Naïve Bayes memiliki kemampuan yang cepat dalam membuat model, mempunyai kemampuan memprediksi dan juga menyediakan metode baru dalam mengeksplor dan memahami data. Algoritma Naïve Bayes hanya mendukung pada atribut yang bertipe data discrete atau discretized, atau tidak mendukung atribut yang bernilai continuous (numerik) dan semua atribut dapat menjadi independen, menjadi atribut yang memberi kontribusi kepada atribut yang diprediksi.

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Bayesian classification didasarkan 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 (Wu *et al.*, 2008).

Teorema Bayes memiliki bentuk umum sebagai berikut:

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

Berdasarkan rumus di atas kejadian H merepresentasikan sebuah kelas dan X merepresentasikan sebuah atribut. $P(H)$ disebut prior probability H , contoh dalam kasus ini adalah probabilitas kelas yang mendeklarasikan normal. $P(X)$ merupakan prior probability X , contoh untuk probabilitas sebuah atribut protocol_type. $P(H|X)$ adalah posterior probability yang merefleksikan

probabilitas munculnya kelas normal terhadap data atribut protocol_type. $P(X|H)$ menunjukkan kemungkinan munculnya prediktor X (protocol_type) pada kelas normal. Dan begitu juga seterusnya untuk proses menghitung probabilitas keempat kelas lainnya.

C. Rapid Miner

Rapidminer adalah aplikasi perangkat lunak yang menyediakan lingkungan yang terintegrasi untuk persiapan data, pembelajaran mesin, penggalian data, penggalian teks, dan menganalisa prediksi. Rapid Miner sebelumnya dikenal dengan nama YALE (Yet Another Learning Environment) dikembangkan mulai tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer dari bagian kecerdasan buatan Universitas Teknik Dortmund. Pengembangan Rapidminer di dorong oleh rapid-I mulai tahun 2006, rapid-I adalah perusahaan berdiri dibawah Ingo Mierswa dan Ralf Klinkenberg . Tahun 2007 namanya berubah dari YALE menjadi Rapidminer. Aplikasi ini untuk tujuan bisnis, penelitian, pendidikan , pelatihan, visualisasi Prototipe, dan pengembangan platform aplikasi serta mendukung semua proses pembelajaran mesin diantaranya preparing data, visualisasi, validitas modeling, dan pengembangan. Rapidminer dikembangkan dengan model core terbuka. Dengan Rapid Studio Free Edition yang dibatasi 1 processor logika dan 10.000 baris data terintegrasi.(Naik and Samant, 2016).

D. Bagging

Bagging ditemukan oleh Breiman (1996) yang merupakan kepanjangan dari " bootstrap aggregating" (Wu and Kumar, 2009). Bagging adalah salah satu teknik dari ensemble method dengan cara memanipulasi data training, data training di duplikasi sebanyak d kali dengan pengembalian (sampling with replacement), yang akan menghasilkan sebanyak d data training yang baru, kemudian dari data d training tersebut akan dibangun classifier-classifier yang disebut sebagai bagged classifier(Altman and Krzywinski, 2017). Karena data pada tehnik bagging dilakukan sampling with replacement, ukuran data bagging sama dengan data aslinya, tetapi

distribusi data dari tiap data bagging berbeda, beberapa data dari data training bisa saja muncul beberapa kali atau mungkin tidak muncul sama sekali (Galar *et al.*, 2012). Hal inilah yang menjadi kunci kenapa bagging bisa meningkatkan akurasi karena dengan sampling with replacement dapat memperkecil variance dari dataset (Mordelet and Vert, 2014)

E. Evaluasi Algoritma Klasifikasi Data Mining

D.1 Evaluasi Confusion Matrix

Untuk melakukan evaluasi terhadap model klasifikasi berdasarkan perhitungan objek testing mana yang diprediksi benar dan tidak benar. Confusion Matrix berisi informasi tentang aktual (actual) dan prediksi (predicted) pada sistem klasifikasi. Kinerja sistem seperti ini biasanya dievaluasi dengan menggunakan data pada matriks. Perhitungan ini ditabulasikan kedalam tabel yang disebut Confusion Matrix (Luque *et al.*, 2019). Bentuk Confusion Matrix dapat dilihat pada Tabel berikut ini:

Tabel 2.1 Confusion Matrix

CLASSIFICATION		PREDICTED CLASS	
		CLASS:YES	CLASS:NO
OBSERVED CLASS	CLASS:YES	a	b
	CLASS:NO	c	d

(True Positive-TP) (False Negative-FN) (False Positive-FP) (True negative-TN)

Pada Tabel 2.1 untuk True positive merupakan tupel positif di data set yang diklasifikasikan positif, True negatives merupakan tupel negatif di data set yang diklasifikasikan negatif. False positives adalah tupel positif di data set yang diklasifikasikan negatif False negatives merupakan jumlah tupel negatif yang diklasifikasikan positif.

Setelah dilakukan confusion matrix berikutnya akan dihitung accuracy, sensitivity, specificity, PPV, NPV. Sensitivity digunakan untuk membandingkan jumlah true positives terhadap jumlah tupel yang positives sedangkan specificity adalah perbandingan jumlah true negatives terhadap jumlah tupel yang negatives. Sedangkan untuk PPV(Positive Predictive Value atau nilai prediktif positif) adalah proporsi kasus

dengan hasil diagnosa positif, NPV(Negative Predictive Value atau nilai prediktif negatif) adalah proporsi kasus dengan hasil diagnosa negatif.

Confusion matrix memberikan penilaian kinerja model klasifikasi berdasarkan jumlah objek yang diprediksi dengan benar dan salah (Gorunescu, 2011). Akurasi kelas minoritas dapat menggunakan metrik TP rate/recall (sensitivitas).

Rumus-rumus yang digunakan untuk melakukan penghitungannya adalah:

Keakuratan (*Accuracy*) adalah proporsi jumlah prediksi yang benar. Hal ini ditentukan dengan menggunakan rumus *accuracy* berikut :

$$Accuracy = \frac{a + b}{a + b + c + d} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

F. D.2 Kurva ROC

Kurva ROC banyak digunakan untuk menilai hasil prediksi, kurva ROC adalah teknik untuk memajemen pengklasifikasian berdasarkan kinerja mereka (Hoo, Candlish and Teare, 2017). Kurva ROC merupakan alat dua dimensi yang digunakan untuk menilai kinerja klasifikasi yang menggunakan dua class keputusan, tiap objek dipetakan ke salah satu elemen dari himpunan pasangan, positif atau negatif. Pada kurva ROC, TP rate diplot pada sumbu Y dan FP rate diplot pada sumbu X. Untuk klasifikasi data mining menurut Gorunescu, nilai AUC dapat dibagi menjadi beberapa kelompok (Ridwan, Andono and Supriyanto, 2018):

1. Nilai 0,90 – 1,00= Excellent Classification
2. Nilai 0,80 – 0,90= Good Classification
3. Nilai 0,70 – 0,80= Fair Classification
4. Nilai 0,60 – 0,70= Poor Classification
5. Nilai 0,50 – 0,60= Failure

III. METODE PENELITIAN

Pendekatan utama dalam penelitian ini yaitu pendekatan kualitatif dan pendekatan kuantitatif. Dalam penelitian ini metode yang digunakan yaitu metode penelitian kuantitatif. Tujuan dari penelitian ini adalah melakukan klasifikasi dan evaluasi model algoritma Naïve Bayes dan Bagging+naïve bayes untuk mengetahui akurasi masing-masing algoritma

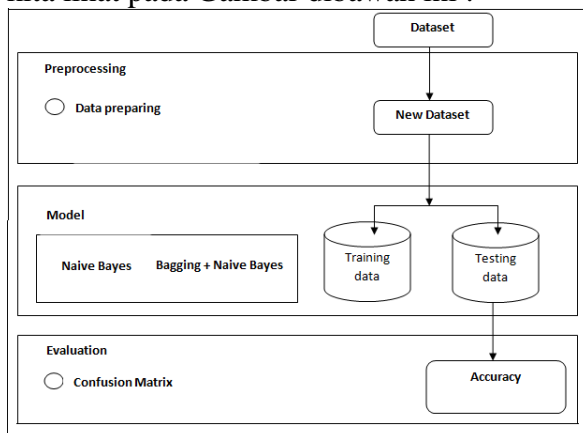
dalam mengklasifikasikan blogger Profesional

A. Sumber Data

Sumber data yang digunakan pada penelitian ini adalah data sekunder dari dataset UCI Machine Learning Repository dengan alamat web <https://archive.ics.uci.edu/ml/machine-learning-databases/00255/>. Dataset yang digunakan adalah Blogger dataset dimana file tersebut bernama kohkiloyeh.xls. Variabel yang digunakan pada penelitian ini adalah sebanyak 6 variabel dengan jumlah data sebanyak 100. Pengumpulan informasi untuk membentuk database dilakukan dengan daftar pertanyaan. Kuesioner ini diberikan secara lisan, tertulis dan juga pemrograman situs web yang menyertakan internet kuesioner dan pengguna dapat menjawab pertanyaan sebagaimana mereka inginkan. Data dikumpulkan dari Provinsi Kohkiloye dan Boyer Ahmad di Iran..

B. Kerangka Pemikiran

Kerangka penelitian dari penelitian bisa kita lihat pada Gambar dibawah ini :



Gambar 3.1 Kerangka Pemikiran

Adapun kerangka pemikiran pada Gambar 3.1 dapat dijelaskan sebagai berikut :

1) Preprocessing

Preprocessing data dilakukan dengan cara menangani nilai yang hilang mengikuti teknik mengabaikan tupel dengan nilai yang tidak lengkap. Setelah praproses, total 100 instans telah tersisa.

2) Model

Tool RapidMiner diterapkan pada Dataset yang baru untuk di Training maupun di Testing pada Algoritma Naïve Bayes dan juga

mentestingnya pada Algoritma Bagging-naive bayes.

3) Evaluasi

Kemudian Dataset diuji/dievaluasi dengan Confusion Matrix serta diukur tingkat akurasi.

IV. HASIL DAN PEMBAHASAN

Variabel data yang digunakan pada penelitian ini disajikan pada Tabel 4.1 yakni sebagai berikut.

Tabel 4.1 Deskripsi Variable Dataset

No.	Atribut	Value
1	Degree	1.High, 2.Medium, 3.Low
2	Caprice	1.Left, 2.middle, 3. Right
3	Topic	1.Impression, 2.political, 3.news, 4. Tourism, 5.Sceintific
4	lmt	1.Yes, 2.No.
5	lpss	1.Yes, 2.No.
6	pb	1.Yes, 2.No.

Dari Tabel 4.1 Ada 5 variabel dataset gejala dan 1 variabel class penentu klasifikasi. Pada dataset ini atributnya terdiri : Pendidikan (degree), tingkah politik(caprice), topik, media local turnover (LMT) dan ruang lokal, politik dan sosial (LPSS) dan penentunya adalah professional blogger (pb).

A. Preprocessing

Dari dataset yang akan di analisis dijadikan sebagai Data Training dan Data Testing yang ada di klasifikasikan oleh algoritma Naïve Bayes dan bagging+naïve bayes

Adapun contoh data training dan data testing dapat di lihat pada tabel 4.2 :

Tabel 4.2 Data Training dan Testing

No.	Atribut	Value	
		1	2
1	Degree	high	high
2	Caprice	left	left
3	Topic	Impression	political
4	lmt	yes	yes
5	lpss	yes	yes
6	pb	yes	yes

Setelah preproses, total 100 data masih valid

B. Model

Setelah Praprocessing, data Training dan Data Testing akan kita proses klasifikasi menggunakan aplikasi Rapidminer Adapun hasil dari Confusion Matrix.nya :

Tabel 4.3 Tabel Hasil Class Recall dan Precision naive bayes

	true yes	true no	class precision
pred. yes	63	18	77.78%
pred. no	5	14	73.68%
class recall	92.65%	43.75%	

Dari Tabel 4.3 didapatkan Class Precision = 73,68%, dan Class Recall: 43,75 %

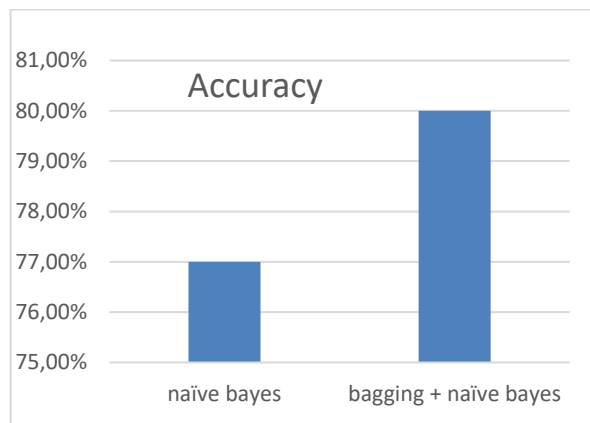
Tabel 4.4 Tabel Hasil Class Recall dan Precision bagging+naive bayes

	true yes	true no	class precision
pred. yes	62	14	81.58%
pred. no	6	18	75.00%

Dari Tabel 4.4 didapatkan Class Precision = 75%, dan Class Recall: 56,25 %

C. Evaluasi

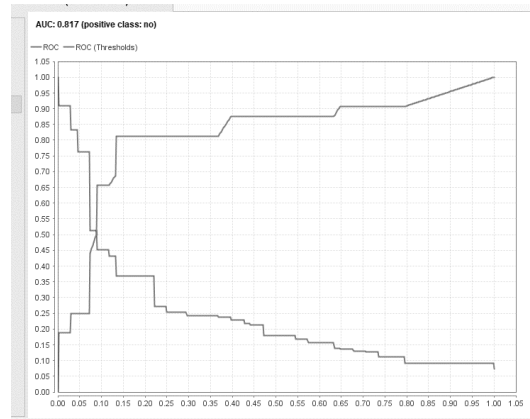
Hasil dari evaluasi pengklasifikasian dengan Naive Bayes dan bagging+naive bayes menghasilkan :



Gambar 4.1 Akurasi Naive Bayes dan bagging+naive bayes

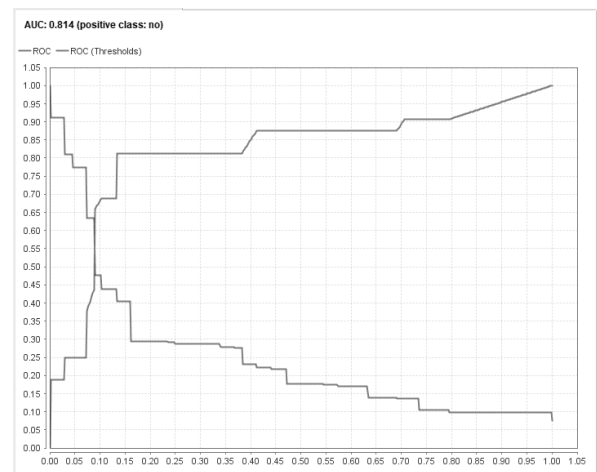
Dari gambar diatas didapatkan akurasi untuk mengklasifikasikan Dataset blogger yaitu dengan naive bayes 77 % sedangkan untuk bagging+naive bayes 80%.

Adapun hasil dari pengujian ini juga menghasilkan gambar 4.2 dibawah ini



Gambar 4.2 Kurva Roc Model Naive Bayes

Kurva ROC (Reciver Operating Characteristic) diatas menunjukkan algoritma Naive Bayes memiliki nilai AUC sebesar 0.817 yang artinya Good Classification (Bagus) Dalam hasil penelitian, menunjukkan Naive Bayes memberikan akurasi yang bagus untuk prediksi blogger professional pada UCI dataset Machine Learning Repository.



Gambar 4.3 Kurva Roc Model bagging+Naive Bayes

Gambar diatas Kurva ROC (Reciver Operating Characteristic) diatas menunjukkan algoritma Naive Bayes memiliki nilai AUC sebesar 0.826 yang artinya Good Classification (Bagus) ini artinya bagging meningkatkan nilai AUC

V. KESIMPULAN

Terdapat 5 atribut yaitu terdiri : Pendidikan (degree), tingkah politik(caprice), topik, media local turnover (LMT) dan ruang lokal,politik dan sosial (LPSS) dan penentunya adalah Professional blogger (pb).

Penelitian ini membandingkan Algoritma Naive Bayes dan bagging+naive Bayes untuk

pengklasifikasian Dataset blogger. Data ini terdiri 100 data.

Dari hasil proses Klasifikasi permodelan algoritma naïve bayes terhadap Dataset Blogger menghasilkan Class Precision = 73,68%, dan Class Recall: 43,75 % sedangkan untuk algoritma bagging+naïve bayes yaitu Class Precision = 75%, dan Class Recall: 56,25 % ini artinya Teknik bagging dapat meningkatkan class recall dan class precision pada klasifikasi naïve bayes

Hasil evaluasi klasifikasi ini menghasilkan akurasi untuk mengklasifikasikan Dataset blogger pada naïve bayes yaitu 77% dan untuk bagging +naïve bayes meningkat yaitu 80%

Kurva ROC (Reciver Operating Characteristic) Klasifikasi permodelan algoritma Naïve Bayes memiliki nilai AUC sebesar 0.817 dan untuk bagging+naïve bayes meningkat sebesar 0.826

DAFTAR PUSTAKA

- Altman, N. and Krzywinski, M. (2017) 'Points of Significance: Ensemble methods: Bagging and random forests', *Nature Methods*. doi: 10.1038/nmeth.4438.
- Boukenze, B. *et al.* (2012) 'Top 10 algorithms in data mining', *Knowledge and Information Systems*. doi: 10.1017/S0269888910000032.
- Galar, M. *et al.* (2012) 'A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches', *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*. doi: 10.1109/TSMCC.2011.2161285.
- Gorunescu, F. (2011) *Data Mining, Soft Computing*. Berlin, Heidelberg: Springer
- Berlin Heidelberg. doi: 10.1007/978-3-642-19721-5.
- Hoo, Z. H., Candlish, J. and Teare, D. (2017) 'What is an ROC curve?', *Emergency Medicine Journal*. doi: 10.1136/emered-2017-206735.
- Larose, D. T. (2006) *Data Mining Methods and Models, Data Mining Methods and Models*. doi: 10.1002/0471756482.
- Luque, A. *et al.* (2019) 'The impact of class imbalance in classification performance metrics based on the binary confusion matrix', *Pattern Recognition*. doi: 10.1016/j.patcog.2019.02.023.
- Mordelet, F. and Vert, J. P. (2014) 'A bagging SVM to learn from positive and unlabeled examples', *Pattern Recognition Letters*. doi: 10.1016/j.patrec.2013.06.010.
- Naik, A. and Samant, L. (2016) 'Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime', *Procedia Computer Science*, 85, pp. 662–668. doi: 10.1016/j.procs.2016.05.251.
- Ridwan, A., Andono, P. N. and Supriyanto, C. (2018) 'Optimasi Klasifikasi Status Gizi Balita Berdasarkan Indeks Antropometri Menggunakan Algoritma Naive', *Teknologi Informasi*.
- Wu, X. *et al.* (2008) 'Top 10 algorithms in data mining', *Knowledge and Information Systems*. doi: 10.1007/s10115-007-0114-2.
- Wu, X. and Kumar, V. (2009) *The Top Ten Algorithms in Data Mining*. Taylor & Francis Group.