

Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naive Bayes

Veronica Retno Sari¹, Feranandah Firdausi², Yufis Azhar³.

^{1,2,3} Program Studi Informatika, Universitas Muhammadiyah Malang

Email: veronicaretnos@gmail.com¹, feranandah10@gmail.com², yufis@umm.ac.id³

(Received: 4 Juli 2020/ Accepted: 14 September 2020/ Published Online: 20 Desember)

Abstrak

Klasifikasi merupakan salah satu teknik yang ada di dalam data mining dan berguna dalam pengelompokan data berdasarkan keterikatan data dengan data sampel. Dataset yang digunakan merupakan dataset kopi yang bersumber dari Database Coffee Quality Institute melalui platform GitHub. Atribut yang terdapat pada dataset tersebut ialah *Aroma, Aftertaste, Flavor, Acidity, Balance, Body, Uniformity, Sweetness, Clean Cup*, dan *Copper points*. Terdapat 3 metode klasifikasi yang digunakan dalam penelitian ini, yaitu *Stochastic Gradient Descent, Random Forest* dan *Naive Bayes*. Penelitian ini bertujuan untuk mengetahui algoritma mana yang paling efektif untuk memprediksi kualitas kopi pada dataset. Kemudian hasil prediksi akan diujikan dengan metode *K-Fold Cross Validation* dan *Area Under the Curve (AUC)*. Hasil penelitian menunjukkan bahwa algoritma *Stochastic Gradient Descent* mendapatkan hasil akurasi paling baik dibandingkan dua metode lainnya dengan nilai akurasi sebesar 98% dan meningkat menjadi 99% dengan pengujian menggunakan metode *K-fold Cross Validation* dan *AUC*.

Kata kunci: Klasifikasi, Kopi Arabika, Perbandingan, Prediksi

Abstract

Classification is one of the techniques that exist in data mining and is useful for grouping a data based on the attachment of the data with the sample data. The dataset that is used in this study is the coffee dataset taken from Dataset Coffee Quality Institute on the GitHub platform. The attributes that contained in the dataset are Aroma, Aftertaste, Flavor, Acidity, Balance, Body, Uniformity, Sweetness, Clean Cup, and Copper points. There are 3 classification methods that are used in this study, Stochastic Gradient Descent, Random Forest and Naive Bayes. The aim of this study is to find out which algorithm is the most effective to predict the coffee quality in the dataset. After that, the prediction results will be tested using K-Fold Cross Validation and Area Under the Curve (AUC) method. The results show that Stochastic Gradient Descent obtained the best accuracy results compared to the other two methods with an accuracy of 98% and increased to 99% after tested using K-fold Cross Validation and AUC method.

Keywords: Classifier, Coffee Arabica, Comparison, Prediction

PENDAHULUAN

Kopi merupakan salah satu produk yang diakui di seluruh dunia sebagai produk yang paling banyak dikonsumsi di pasar internasional. Lebih dari 60 spesies biji kopi yang diperdagangkan, dua yang paling populer adalah kopi arabika dan kopi robusta (Santos et al., 2012). Kopi arabika memiliki keunggulan dalam sektor perkebunan kopi, karena keunggulannya dalam memenuhi pangsa pasar dengan minat yang tinggi maka kopi arabika dapat ditemui di Indonesia maupun di luar negeri (Chandra, Ismono, & Kasymir, 2013).

Penilaian kualitas kopi merupakan kunci utama dalam penetapan harga dan menentukan potensi ekspor di negara-negara penghasil kopi. Saat ini, analisis fisik dan *cup quality* merupakan alat utama untuk penilaian kualitas kopi. Kualitas fisik dapat dilihat dari

warna, ukuran dan bentuk biji kopi. Sedangkan *cup quality* dapat dinilai dari rasa, aroma, keasaman, bentuk serta kebersihan cangkirnya (Tolessa, Rademaker, De Baets, & Boeckx, 2016). Sifat fisik dan kimia yang dihasilkan kopi arabika berasal dari beberapa tahapan proses yaitu (1) pengeringan (2) sortirisasi dan grading (3) sangrai (4) penggilingan (Edowai & Tahoba, 2018). Oleh karena itu, dengan adanya beragam proses pemanggangan pada setiap negara-negara penghasil kopi, menghasilkan beragam kualitas kopi pula.

Prediksi kualitas kopi diperlukan untuk mengetahui kualitas kopi terbaik yang dapat digunakan sebagai acuan dalam memilih kopi. Algoritma yang terdapat pada data mining dapat dimanfaatkan untuk memprediksi kualitas kopi. Namun, tidak semua metode pada data mining dapat menghasilkan kinerja model yang baik (Arifin & Sasongko, 2018). Untuk itu, perlu dilakukan pengujian menggunakan beberapa algoritma pada data mining untuk mengetahui algoritma mana yang dapat menghasilkan kinerja terbaik.

Sejumlah peneliti telah menerapkan beberapa metode atau algoritma data mining pada bidang yang berbeda-beda diantaranya pemodelan kerentanan longsor yang dilakukan berbasis GIS (*Geographic Information System*), *Kernel Logistic Regression*, *Naive Bayes Tree* dan *Decision Tree* untuk menganalisis kerentanan longsor yang ada di daerah Taibai (China) (Chen et al., 2017). Kemudian terdapat penelitian yang serupa mengenai analisis kerentanan longsor dan pemetaan di daerah Mu Cang Tai, terletak di Utara Vietnam dengan menggunakan model *Novel Hibryd Naive Bayes Tree* berbasis *Bagging* (Pham & Prakash, 2019).

Jika pada penelitian lain membahas tentang kerentanan longsor dan pemetaan di suatu daerah, berbeda hal dengan penelitian berikut yang digunakan untuk memperkirakan potensi terkena penyakit kanker payudara pada wanita dengan menggunakan ANN dan *Naive bayes* (Saritas & Yasar, 2019). Adapun penelitian mengenai data mining yang lainnya yaitu membahas tentang pengklasifikasian teks dalam pengolahan informasi dengan menggunakan metode *Naive Bayesian* agar secara otomatis dapat melakukan klasifikasi pada dokumen yang dituju (Xu, 2018).

Penelitian ini terinspirasi dari penelitian sebelumnya tentang minuman wine yang di prediksi dengan menggunakan algoritma SGD, *Random Forest*, *Support Vector Machine*, dan *Logistic Regression*, dimana wine memiliki kandungan yang berbeda-beda karena proses pembuatan yang beragam sehingga menghasilkan kualitas yang berbeda pula. Penelitian tersebut mendapatkan hasil kinerja algoritma terbaik dari algoritma *Random Forest* dengan akurasi 87,33% (Pawar, Mahajan, & Bhoithe, 2019).

Pada penelitian sebelumnya dalam kasus kualitas dan keaslian makanan, menyatakan bahwa *support vector machine* (SVM) menghasilkan tingkat kinerja terbaik dibandingkan dengan metode *random forest* (RF) dan *classification and regression tree* (CART) (Jiménez-Carvelo, González-Casado, Bagur-González, & Cuadros-Rodríguez, 2019). Pada penelitian perbandingan teknik klasifikasi dalam data mining untuk bank direct marketing menyatakan bahwa model *Stochastic Gradient Descent* memiliki nilai akurasi terbaik yaitu sebesar 0,972 dibandingkan dengan 9 teknik klasifikasi lainnya (Oktanisa & Supianto, 2018).

Perbedaan algoritma *Stochastic Gradient Descent* atau SGD dibandingkan dengan metode yang lain. Algoritma *Stochastic Gradient Descent* merupakan model dalam *deep learning* yang mengoptimasi fungsi dengan mengikuti gradient yang memiliki nosy dengan ukuran langkah yang menurun (Mandt, Hof Fman, & Blei, 2017). Algoritma *Random Forest* merupakan pengembangan dari decision tree, setiap decision tree akan melakukan training data menggunakan sampel individu dan setiap atribut dipecah menjadi tree yang akan dipilih antara atribut subset yang bersifat acak (Budi Adnyana, 2016).

Algoritma Naive Bayes merupakan metode probabilistik yang berdasar pada teorema Bayes. Algoritma ini tergolong cepat dan sederhana dengan mengasumsikan fitur yang dapat berdiri sendiri (Oktanisa & Supianto, 2018). Maka dari itu, pada penelitian ini menggunakan

dataset kopi arabika yang mana memiliki kualitas berbeda-beda akibat dari beragamnya proses pemanggangan yang dilakukan oleh setiap negara penghasil kopi dengan menggunakan beberapa algoritma.

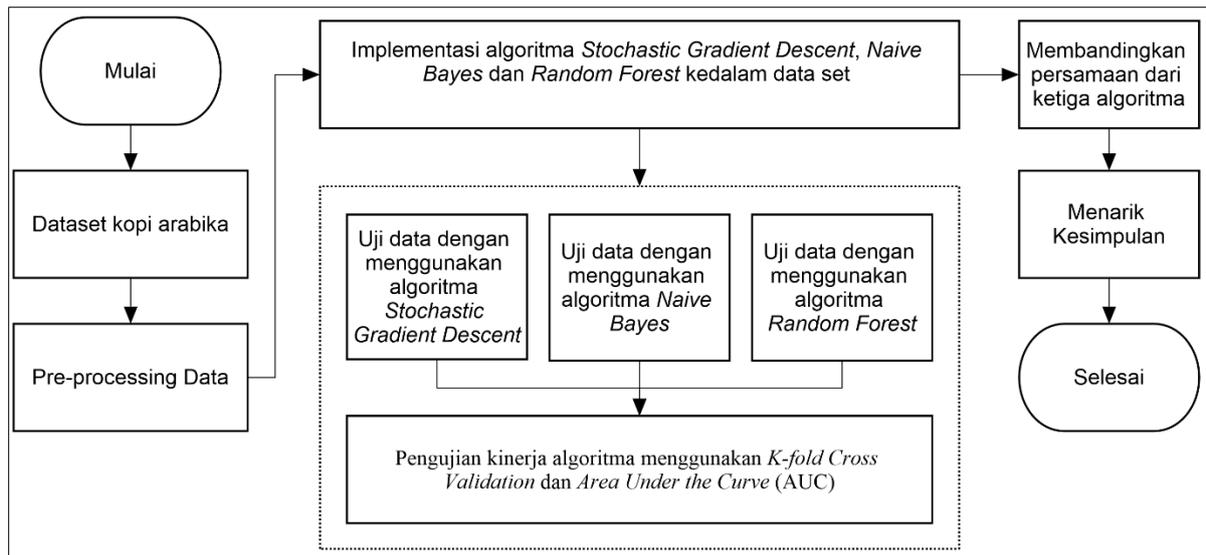
Penelitian kualitas kopi menggunakan algoritma *Stochastic gradient descent*, *Naive bayes*, dan *Random forest* ini digunakan untuk membandingkan efisiensi data dengan algoritma yang digunakan. Hasil pada masing-masing algoritma akan dibandingkan berdasarkan akurasi dari setiap algoritma. Pengukuran kinerja algoritma dalam penelitian ini menggunakan *F1-score*. *F1-score* merupakan gabungan dari 2 metrik yaitu : *Precision* yang menyatakan seberapa akurat algoritma yang digunakan dalam memprediksi kelas dan *Recall* yang menyatakan kebalikan dari kegagalan, karena hilangnya suatu kejadian yang salah dalam pengklasifikasian (Pawar et al., 2019).

Kemudian hasil akurasi yang memiliki nilai tertinggi, membuktikan bahwa algoritma tersebut mampu memprediksi kualitas kopi arabika dengan baik. Setelah hasil akurasi dari masing-masing algoritma telah ditentukan, maka perlu adanya pengujian untuk menilai proses kinerja algoritma yang digunakan. *K-fold Cross Validation* digunakan untuk melakukan pengujian, dengan membagi data sampel secara sebarang kemudian data tersebut dikelompokkan sebanyak nilai K *k-fold* (Anjelika, 2018).

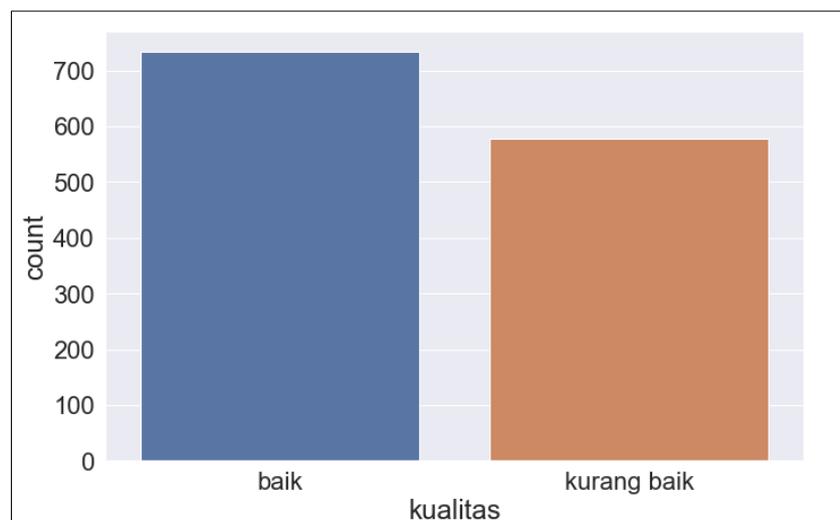
K-fold adalah metode yang biasa digunakan untuk menguji sistem agar dapat diketahui rata-rata keberhasilan dari suatu sistem. Proses *K-fold* melakukan perulangan dengan mengacak atribut masukan kemudian sistem tersebut akan teruji untuk beberapa atribut input yang acak (Pitria, 2014). Selain itu, untuk memastikan lagi kinerja algoritma, maka dilakukan pengujian kembali menggunakan *Area Under the Curve* (AUC). AUC adalah suatu daerah di bawah *Receiver operating character* (ROC). Nilai *Area Under the Curve* (AUC) secara teoritis berada di antara 0 dan 1. Memberikan pandangan tentang keseluruhan pengukuran terhadap kesesuaian dari model yang digunakan. Semakin besar *area under the curve* maka semakin baik variabel yang diteliti dalam memprediksi kejadian (Maskoen & Purnama, 2018). Nilai AUC digunakan untuk menentukan hasil klasifikasi (Karyadiputra & Hijriana, 2018). Nilai AUC dalam klasifikasi data mining, dibagi kedalam beberapa kelompok yaitu, klasifikasi sangat baik = 0.90 - 1.00, klasifikasi baik = 0.80 – 0.90, klasifikasi cukup = 0.70 – 0.80, klasifikasi buruk = 0.60 – 0.70 dan klasifikasi salah = 0.50 – 0.60 (Gorunescu, 2011). Tujuan penelitian ini adalah membandingkan tingkat kinerja algoritma yang dihasilkan dari algoritma *Stochastic gradient descent*, *Random forest*, dan *Naive bayes* dalam memprediksi kualitas kopi arabika.

METODE

Tahap penelitian dalam mengimplementasikan algoritma *Stochastic Gradient Descent*, *Naive Bayes*, dan *Random Forest* pada dataset, dapat dilihat pada gambar 1. Penelitian menggunakan bahasa pemrograman python sebagai media dalam membuat perbandingan prediksi pada kualitas kopi arabika. Pemilihan bahasa pemrograman disini cukup penting untuk dilakukan, supaya program yang akan dikembangkan dapat bekerja secara optimal (Takdirillah, 2020). Dataset yang digunakan diperoleh dari GitHub (<https://github.com/jldbc/coffee-quality-database>), dengan jumlah data sebanyak 1313 data dan dikelompokkan menjadi 2 kelas (kualitas kopi baik dan kurang baik). Plot bar di bawah ini, menunjukkan jumlah data dari pengelompokan kualitas kopi yang baik dan kurang baik, dengan jumlah kopi kualitas baik sebanyak 734 dan kopi kurang baik sebanyak 578.



Gambar 1. Tahap Implementasi Algoritma pada Dataset



Gambar 2. Plot Bar Pengelompokan Kualitas Kopi

Pre-processing data dilakukan dengan memilih variabel independen dan dependen, mengganti tipe data sesuai dengan kebutuhan dan membagi data menjadi data training dan data testing. Setelah dilakukan *pre-processing* data, didapatkan dataset yang terdiri dari 10 variabel independen (*Aroma, Aftertaste, Flavor, Acidity, Balance, Body, Uniformity, Sweetness, Clean Cup, dan Copper points*) dan 1 variabel dependen (*quality_score*). Kemudian dataset tersebut dibagi menjadi 20% data testing dan 80% data training. Data training digunakan untuk mendapatkan nilai pada tabel probabilitas dan data testing digunakan untuk menguji tabel probabilitas yang dihasilkan oleh data training (Ridwan, Suyono, & Sarosa, 2013).

Setelah itu dilanjutkan dengan implementasi algoritma *Stochastic Gradient Descent, Naive Bayes, dan Random Forest* ke dalam dataset. Metode *Stochastic Gradient Descent* (SGD) merupakan salah satu metode yang memiliki optimasi iteratif dalam menemukan titik yang meminimumkan suatu fungsi yang dapat diturunkan. Metode SGD mulai bekerja dari tebakan awal kemudian secara iteratif tebakan itu didapat lalu diperbaiki, hal tersebut

berdasarkan suatu aturan yang melibatkan gradien atau turunan pertama dari fungsi yang ingin diminimumkan (Umar, Riadi, & Purwono, 2020).

Naive Bayes adalah klasifikasi statistik, biasanya digunakan untuk memprediksi probabilitas keanggotaan suatu *class* (Januarsyah, Zuhairi, & Malik, 2019). Kemudian untuk metode RF atau random esembles yang berarti hutan acak merupakan kombinasi pohon keputusan sehingga setiap pohon bergantung terhadap nilai - nilai vektor acak yang di sampling secara independen dengan distribusi yang sama untuk semua pohon tersebut. Keunggulan metode ini, dapat memproses pemilihan fitur yang digunakan secara acak kemudian memilah setiap simpul agar menghasilkan tingkat kesalahan atau error yang relatif rendah (Umar et al., 2020).

Pengukuran kinerja algoritma dilakukan untuk menentukan algoritma mana yang paling optimal. *F1-score* akan digunakan untuk mengukur kinerja dari masing - masing algoritma. Setelah mengetahui akurasi dari masing-masing algoritma, maka dilakukan pengujian kinerja model pada dataset menggunakan *K-fold Cross Validation* yang bertugas untuk menguji kinerja klasifikasi sistem dengan jumlah *fold* = 10. Untuk memastikan lagi keakuratan dataset maka dilakukan pengujian dengan menggunakan *Area Under the Curve* (AUC). Tahapan selanjutnya, melakukan analisa perbandingan dari setiap algoritma yang digunakan. Kemudian dilanjutkan dengan menarik kesimpulan algoritma mana yang memiliki hasil terbaik dari masing-masing algoritma tersebut.

HASIL DAN PEMBAHASAN

Hasil

Penelitian yang dilakukan dengan menggunakan perbandingan algoritma, didapatkan hasil akurasi yang berbeda-beda dari setiap model yang digunakan. Adapun hasil akurasinya yang terlihat pada tabel 1. Algoritma *Stochastic Gradient Descent* memberikan tingkat akurasi sebesar 99%, algoritma *Naive Bayes* memberikan tingkat akurasi sebesar 79%, dan algoritma *Random Forest* memberikan tingkat akurasi sebesar 96% yang artinya tingkat akurasi tertinggi ada dimetode *Stochastic Gradient Descent*. Algoritma tersebut juga diuji dengan kurva pengujian AUC untuk melihat kinerja model mengalami penurunan atau tidak.

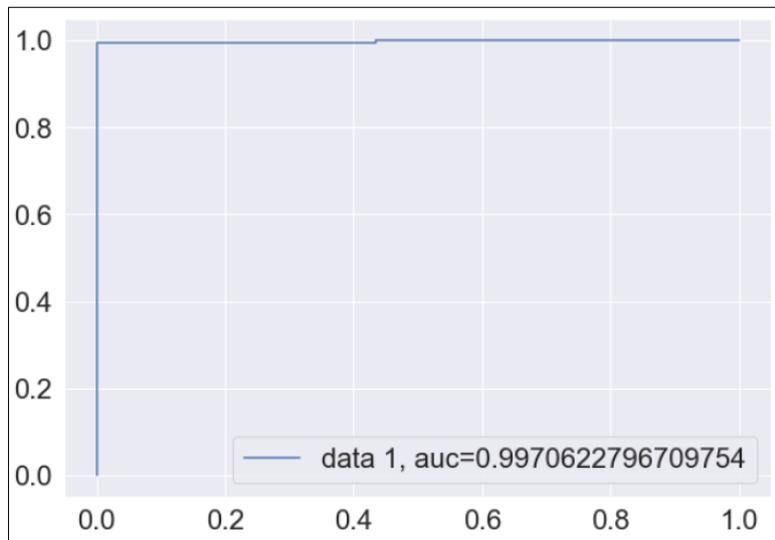
Tabel 1. Performa Algoritma Data Mining dalam Pengklasifikasian Kualitas Kopi Arabika

Metode	Kelas	Precision	Recall	F1-Score	Acuracy	Cross Validation	AUC
Stochastic Gradient Descent	Baik	0.99	0.99	0.99			
	Kurang Baik	0.99	0.98	0.99	99%	99%	99%
Naive Bayes	Baik	0.76	0.92	0.83			
	Kurang Baik	0.86	0.63	0.73	79%	78%	94%
Random Forest	Baik	0.94	0.99	0.96			
	Kurang Baik	0.99	0.93	0.95	96%	95%	99%

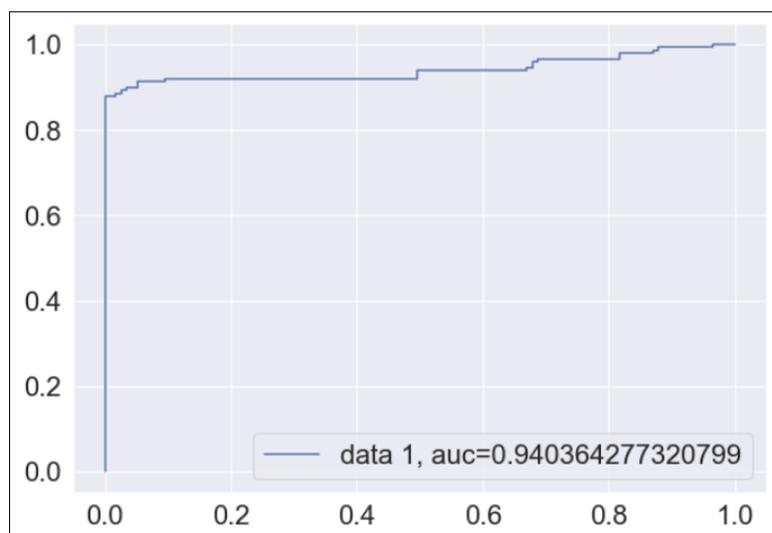
Pada algoritma *Stochastic Gradient Descent* kinerja model mencapai 99% sehingga pada pengujian tersebut algoritma yang digunakan tidak mengalami penurunan kinerja model, begitu juga dengan pengujian dengan menggunakan *Cross Validation* kinerja modelpun tidak mengalami penurunan, berbeda dengan algoritma *Naive bayes* dan *Random Forest*. Pada algoritma *Naive bayes* kinerja model pada pengujian *Cross Validation* hanya 78% dibandingkan algoritma yang lain dan mengalami kenaikan kinerja dengan menggunakan kurva AUC hingga 94% yang artinya hasil kinerja model baik untuk digunakan pada dataset kopi, akan tetapi hasil tersebut memiliki perbandingan jauh dengan metode

Stochastic Gradient Descent yang memiliki tingkat pengujian yang konsisten dan lebih tinggi kinerja modelnya.

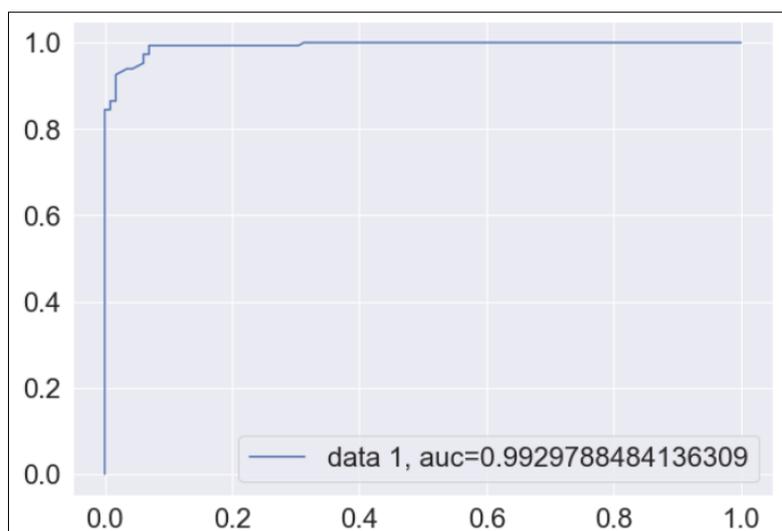
Begitu juga pada algoritma *Random Forest* yang mengalami kinerja model 95% pada pengujian *Cross Validation* sementara dengan pengujian AUC mencapai 99%, Hasil pengujian *Random Forest* dengan *Stochastic Gradient Descent* memang memiliki hasil yang sama yaitu 99% akan tetapi dari pengujian *Cross Validation* algoritma *Random Forest* tidak mengalami hasil yang konsisten dan tingkat akurasi yang ada masih lebih baik algoritma *Stochastic Gradient Descent*. Berikut kurva pengujian AUC yang nampak pada gambar 3, 4, dan 5.



Gambar 3. Kurva AUC pada Algoritma SGD



Gambar 4. Kurva AUC pada Algoritma Naive Bayes



Gambar 5. Kurva AUC pada Algoritma Random Forest

Pada gambar 3, 4 dan 5 merupakan hasil kinerja model dengan pengujian AUC, pada kinerja model algoritma *Stochastic Gradient Descent* dapat dilihat pada gambar 3. Pengujian AUC pada algoritma ini menghasilkan tingkat kinerja model sebesar 99%. Sementara pada kinerja model algoritma *Naïve Bayes* dapat dilihat pada gambar 4. Pengujian AUC pada algoritma ini menghasilkan tingkat kinerja model sebesar 94%. Dan kinerja model algoritma *Random Forest* dapat dilihat pada gambar 5. Pengujian AUC pada algoritma ini menghasilkan tingkat kinerja model sebesar 99%

Pembahasan

Dataset yang digunakan merupakan dataset kualitas kopi arabika. Dari dataset ini dilakukan pengujian prediksi dengan menggunakan tiga algoritma yaitu *Stochastic Gradient Descent*, *Random Forest*, dan *Naive Bayes*. Dalam proses pengujiannya, sebanyak 1313 data digunakan, dengan kelas “Baik” sebanyak 734 data dan “Kurang Baik” sebanyak 578 data. Data tersebut dibagi kedalam 2 bagian yaitu data testing dan data training. Hasil pengujian masing-masing algoritma dapat dilihat pada Tabel 1. Nilai yang dihasilkan dari perbandingan algoritma ini, antara lain seperti *precision*, *recall*, *F1-score*, dan *accuracy* serta pengujian kinerja model.

Nilai *precision* tertinggi dalam kelas “Baik” yaitu 99% oleh algoritma *Stochastic Gradient Descent*, sedangkan nilai terendahnya yaitu 76% oleh algoritma *Naive Bayes* dan nilai *precision* tertinggi dalam kelas “Kurang Baik” yaitu 99% oleh algoritma *Random Forest* dan *Stochastic Gradient Descent*, sedangkan nilai terendahnya yaitu 86% oleh algoritma *Naive Bayes*. Nilai *recall* tertinggi dalam kelas “Baik” yaitu 99% oleh algoritma *Stochastic Gradient Descent* dan *Random Forest*, sedangkan nilai terendahnya yaitu 92% oleh algoritma *Naive Bayes* dan nilai *recall* tertinggi dalam kelas “Kurang Baik” yaitu 98% oleh algoritma *Stochastic Gradient Descent*, sedangkan nilai terendahnya yaitu 63% oleh algoritma *Naive Bayes*. Nilai *accuracy* tertinggi yaitu 99% oleh algoritma *Stochastic Gradient Descent*, sedangkan nilai *accuracy* terendah yaitu 79% oleh algoritma *Naive Bayes*.

Hasil pengujian kinerja model dengan *Cross Validation*, didapatkan nilai tertinggi yaitu 99% oleh algoritma *Stochastic Gradient Descent*, sedangkan nilai terendahnya yaitu 78% oleh algoritma *Naive Bayes*. Untuk hasil pengujian kinerja dengan AUC dapat dilihat pada Gambar 3, 4 dan 5 dengan nilai rata-rata dari ketiga algoritma adalah 97.3%. Secara keseluruhan, dengan hasil yang telah didapatkan pada tabel 1, maka dapat dilihat bahwa algoritma *Stochastic Gradient Descent* merupakan algoritma yang paling optimal, dengan menghasilkan tingkat akurasi sebesar 99% serta kinerja model dari *Cross Validation* dan

AUC sebesar 99% dibandingkan dengan algoritma *random forest* dan *naïve bayes*. Berdasarkan dari pengelompokan AUC, maka dapat ditarik kesimpulan bahwa algoritma *Stochastic Gradient Descent* termasuk kedalam klasifikasi sangat baik dengan menghasilkan nilai AUC diantara 0.90 - 1.00, yaitu sebesar 0.997.

Sementara itu, pada penelitian sebelumnya dengan kasus kualitas wine, algoritma *Stochastic Gradient Descent* memberikan tingkat akurasi sebesar 81% dan algoritma *Random Forest* memberikan tingkat akurasi sebesar 87,33% (Pawar et al., 2019). Hal ini membuktikan bahwa tidak semua algoritma pada data mining memberikan kinerja model yang baik dalam mengklasifikasi dataset. Sehingga perlu adanya pengujian pada dataset, dengan menggunakan beberapa algoritma untuk mengetahui algoritma mana yang menghasilkan kinerja terbaik.

SIMPULAN

Berdasarkan hasil penelitian, secara keseluruhan algoritma *Stochastic Gradient Descent* menghasilkan kinerja yang paling baik dan termasuk kategori klasifikasi sangat baik, dibandingkan dengan algoritma *Random Forest* dan *Naive Bayes*. Dilihat dari tingkat akurasi yang dihasilkan oleh algoritma *Stochastic Gradient Descent* sebesar 98% dengan pengukuran kinerja model menggunakan *k-fold cross validation* yang menghasilkan tingkat kinerja model sebesar 99% serta AUC 99%. Dikarenakan penelitian perbandingan algoritma dengan menggunakan *Stochastic Gradient Descent*, *Naive Bayes* dan *Random Forest* untuk dataset kualitas kopi baru dilakukan, maka peneliti berharap dapat diteliti lebih lanjut dengan menggunakan perbandingan teknik pengklasifikasian yang lain ataupun dengan menggunakan teknik pengelompokan dengan harapan dapat menemukan teknik ataupun metode yang lebih efisien dan menghasilkan nilai prediksi yang lebih baik lagi.

REFERENSI

- Anjelika, T. & I. (2018). Penerapan Algoritma Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3957–3961.
- Arifin, O., & Sasongko, T. B. (2018). Analisa Perbandingan Tingkat Performansi Metode Support Vector Machine dan Naive Bayes Classifier Untuk Klasifikasi Jalur Minat SMA. *Seminar Nasional Teknologi Informasi Dan Multimedia 2018*, 6(1), 67–72.
- Budi Adnyana, I. M. (2016). Prediksi Lama Studi Mahasiswa Dengan Metode Random Forest (Studi Kasus: Stikom Bali). *CSRID (Computer Science Research and Its Development Journal)*, 8(3), 201–208. <https://doi.org/10.22303/csrid.8.3.2016.201-208>
- Chandra, D., Ismono, R. H. dan, & Kasymir, E. (2013). Prospek Perdagangan Kopi Robusta Indonesia di Pasar Internasional. *JIIA Jurnal Ilmu Ilmu Agribisnis*, 1(1), 10–15.
- Chen, W., Xie, X., Peng, J., Wang, J., Duan, Z., & Hong, H. (2017). GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models. *Geomatics, Natural Hazards and Risk*, 8(2), 950–973. <https://doi.org/10.1080/19475705.2017.1289250>
- Edowai, D. N., & Tahoba, A. E. (2018). Proses produksi dan uji mutu bubuk kopi arabika (*coffea arabica* L) asal kabupaten Dogiyai, Papua. *Agriovet*, 1(1), 1–18.
- Gorunescu, F. (2011). *Data Mining - Concepts, Models and Techniques*. <https://doi.org/10.16309/j.cnki.issn.1007-1776.2003.03.004>
- Harismawan, A. F., Kharisma, A. P., & Afirianto, T. (2018). Analisis Perbandingan Performa Web Service Menggunakan Bahasa Pemrograman Python, PHP, dan Perl pada Client Berbasis Android. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(2), 237–245.
- Januarsyah, M. F., Zuhairi, E., & Malik, R. F. (2020). Perbandingan Algoritma Random

- Forest, Decision Stump, Naïve Bayes, Bayesian Network dan Algoritma C4. 5 Untuk Prediksi Pola Kartu Poker. *Annual Research Seminar (ARS)*, 5 (1), 122-126.
- Jiménez-Carvelo, A. M., González-Casado, A., Bagur-González, M. G., & Cuadros-Rodríguez, L. (2019). Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity—A review. *Food research international*, 122, 25-39. <https://doi.org/10.1016/j.foodres.2019.03.063>
- Karyadiputra, E., & Hijriana, N. (2018). Analisis Penerapan Algoritma Naive Bayes Untuk Klasifikasi Prioritas Pengembangan Jalan Di Provinsi Kalimantan Selatan. *Technologia: Jurnal Ilmiah*, 9(2), 105-108. <https://doi.org/10.31602/tji.v9i2.1374>
- Mandt, S., Hoffman, M. D., & Blei, D. M. (2017). Stochastic gradient descent as approximate bayesian inference. *The Journal of Machine Learning Research*, 18(1), 4873-4907.
- Maskoen, T. T., & Purnama, D. (2018). Area Under the Curve dan Akurasi Cystatin C untuk Diagnosis Acute Kidney Injury pada Pasien Politrauma. *Majalah Kedokteran Bandung*, 50(4), 259–264. <https://doi.org/10.15395/mkb.v50n4.1342>
- Oktanisa, I., & Supianto, A. A. (2018). Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank a Comparison of Classification Techniques in Data Mining for. *Teknologi Informasi Dan Ilmu Komputer*, 5(5), 567–576. <https://doi.org/10.25126/jtiik20185958>
- Pawar, D., Mahajan, A., & Bhoithe, S. (2019). Wine Quality Prediction using Machine Learning Algorithms. *International Journal of Computer Applications Technology and Research*, 8(9), 385–388. <https://doi.org/10.7753/ijcatr0809.1010>
- Pham, B. T., & Prakash, I. (2019). A novel hybrid model of Bagging-based Naïve Bayes Trees for landslide susceptibility assessment. *Bulletin of Engineering Geology and the Environment*, 78(3), 1911–1925. <https://doi.org/10.1007/s10064-017-1202-5>
- Pitria, P. (2014). Pengguna Twitter Pada Akun Resmi Samsung Indonesia Dengan Menggunakan Naïve Bayes. *Skripsi*. Fakultas Teknologi Industri. Universitas Atma Jaya Yogyakarta.
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Eeccis*, 7(1), 59–64. <https://doi.org/10.1038/hdy.2009.180>
- Santos, K. M., Moura, M. F. V., Azevedo, F. G., Lima, K. M. G., Raimundo, I. M., & Pasquini, C. (2012). Classification of Brazilian Coffee Using Near-Infrared Spectroscopy and Multivariate Calibration. *Analytical Letters*, 45(7), 774–781. <https://doi.org/10.1080/00032719.2011.653905>
- Saritas, M. M., & Yasar, A. (2019). Performance analysis of ANN and Naive Bayes classification algorithm for data classification. *International Journal of Intelligent Systems and Applications in Engineering*, 7(2), 88–91.
- Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan. *Edumatic: Jurnal Pendidikan Informatika*, 4(1), 37–46.
- Tolessa, K., Rademaker, M., De Baets, B., & Boeckx, P. (2016). Prediction of specialty coffee cup quality based on near infrared spectra of green coffee beans. *Talanta*, 150, 367-374. <https://doi.org/10.1016/j.talanta.2015.12.039>
- Umar, R., Riadi, I., & Purwono. (2020). Perbandingan Metode SVM, RF dan SGD untuk Penentuan Model Klasifikasi Kinerja Programmer pada Aktivitas Media Sosial. *Jurnal Resti*, 4(2), 329–335. <https://doi.org/10.29207/resti.v4i2.1770>
- Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science*, 44(1), 48–59.