

KLASIFIKASI VARIETAS KOPI ARABIKA MENGGUNAKAN METODE K-NEAREST NEIGHBOUR

Sufriadi , Bustami , dan Yesy Afrillia

Program Studi Teknik Informatika, Universitas Malikussaleh, Aceh, Indonesia
Email : sufriadi.170170105@mhs.unimal.ac.id

Abstrak

Penelitian ini bertujuan untuk mengklasifikasi varietas kopi arabika dengan menggunakan metode *K-Nearest Neighbor*. Data yang diambil dalam penelitian ini adalah data karakteristik morfologi dari Instansi Penelitian dan Pengkajian Teknologi Pertanian Gayo. Berdasarkan hasil dari pembahasan dan pengujian sistem, maka didapatkan kesimpulan bahwa penelitian ini berhasil membangun sebuah sistem klasifikasi varietas kopi arabika menggunakan metode *K-Nearest Neighbor* yang digunakan untuk menentukan varietas kopi dengan menghitung karakteristik morfologi varietas kopi arabika seperti ukuran buah, cabang yang produktif, bunga, daun, dan batang. Sistem memberikan hasil klasifikasi yaitu sangat unggul, unggul, dan kurang unggul. Unjuk kerja sistem klasifikasi varietas kopi arabika dengan algoritma *K-Nearest Neighbor* dari 54 data pengujian dengan nilai $k=3$ mendapatkan hasil 20 data sebagai kelas sangat unggul, 30 unggul dan 4 kurang unggul. Algoritma *K-Nearest Neighbor* diterapkan dalam klasifikasi varietas kopi arabika mendapatkan tingkat keakurasian dengan $k=3$ sebesar 70%.

Kata Kunci : *Klasifikasi, Varietas, K-Nearest Neighbor.*

Pendahuluan

Sebagai salah satu daerah penghasil kopi arabika provinsi Aceh khususnya di kabupaten Bener Meriah, sebagian besar masyarakatnya menggantungkan hidup sebagai petani kopi. Perkebunan kopi milik rakyat yang ada di Kabupaten Bener Meriah tercatat seluas 43.565 Ha dan produktifitas sebanyak 29.357 ton pada tahun 2018 (BPS Kab. Bener Meriah). Untuk memenuhi kebutuhan kopi yang terus meningkat baik nasional maupun dunia maka diperlukan upaya untuk meningkatkan jumlah produksi kopi yaitu dengan adanya bibit yang berkualitas atau varietas unggul. Varietas kopi Arabika yang dibudidayakan oleh petani di Bener Meriah sangat variatif yang terdiri dari bermacam-macam varietas, bahkan banyak dari varietas-varietas tersebut dicampur dalam satu lahan sempit yang sama. Akibatnya banyak dijumpai varietas-varietas kombinasi baru, yang mempunyai kualitas fisik, hasil produksi dan cita rasa yang berbeda dari varietas aslinya, hal ini merupakan masalah bagi para petani. Pada setiap varietas memiliki ciri yang berbeda dan umumnya petani kopi akan mengganti tanaman kopinya setiap beberapa tahun. Hal ini disebabkan karena menurunnya hasil panen karena umur kopi yang sudah tua dan tidak lagi produktif, oleh karena itu agar petani kopi dapat memilih kopi sesuai dengan yang diinginkan, maka diperlukan sebuah system untuk membedakan varietas kopi yang akan ditanam agar tidak mengurangi kualitas, cita rasa dan agar produksi kopi tetap terjaga kestabilannya. Salah satu cara yang dapat dilakukan untuk menentukan varietas kopi arabika yaitu dengan menggunakan metode klasifikasi berdasarkan karakteristik morfologi.

Penelitian terdahulu tentang klasifikasi yang dilakukan oleh Sistrina Hijrah Wardani yang berjudul Aplikasi Klasifikasi Jenis Tumbuhan Mangrove Berdasarkan Karakteristik Morfologi Menggunakan Metode K-Nearest Neighbor (KNN) Berbasis Web

menghasilkan akurasi yaitu sebesar 77,7% [5]. Penelitian tentang varietas kopi pernah dilakukan oleh Anastasia Novia Widrawati membahas klasifikasi varietas kopi arabika menggunakan metode Support Vector Machine (SVM), penelitian tersebut berguna untuk mengklasifikasikan varietas kopi. Parameter yang digunakan adalah karakteristik uji cita rasa kopi. Akurasi yang didapatkan sebesar 48.33% [6].

Adapun untuk klasifikasi varietas kopi arabika pada penelitian ini penulis menggunakan metode K-Nearest Neighbour. Metode *K-Nearest Neighbour* merupakan suatu metode yang menggunakan algoritma *supervised*. Dimana hasil dari sampel uji baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN yang bertujuan untuk mengklasifikasi objek baru berdasarkan atribut dan training sample. Berdasarkan uraian di atas maka penulis melakukan penelitian ini dengan judul “Klasifikasi Varietas Kopi Arabika Menggunakan Metode K-Nearest Neighbour (KNN)”, dimana parameter yang digunakan dalam penelitian ini yaitu karakteristik morfologis dari kopi seperti buah, bunga, daun dan jenis cabang.

Tinjauan Pustaka

Kopi salah satu jenis tanaman dari sektor perkebunan yang telah lama dibudidayakan dan memiliki nilai ekonomis adalah yang lumayan tinggi. Kopi berasal dari daerah pegunungan Ethiopia di Afrika. Namun, diluar daerah asalnya, yaitu arab tepatnya dinegara yaman kopi baru dikenal oleh masyarakat dunia setelah tanaman tersebut dikembangkan[2]

Data mining adalah proses yang menggunakan Teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Data mining merupakan salah satu cabang ilmu komputer baru[4].

Klasifikasi (*classification*) adalah proses untuk menemukan fungsi atau model yang bisa membedakan kelas data atau konsep yang bertujuan untuk memperkirakan kelas dari objek yang labelnya belum diketahui[3].

K-Nearest Neighbour merupakan suatu metode yang dapat mengklasifikasi berdasarkan data training paling dekat dengan data yang dimiliki sebagai sample untuk menemukan hasil akhir. Metode *K-Nearest Neighbor* adalah metode melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan training sample[1]. Rumus euclidian distance di tunjukkan pada persamaan 2.1

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2} \dots\dots\dots(1)$$

Keterangan:

- Dis = Jarak
- X₁ = Data Latih
- X₂ = Data Uji
- n = Dimensi Data
- i = Variabel Data

Metodologi Penelitian

Sebelum melakukan penelitian, perlu melakukan studi literatur yang berguna untuk mengetahui dasarnya. Studi literatur dilakukan dengan cara mengumpulkan, membaca dan memahami referensi yang terkait dengan klasifikasi varietas kopi dengan metode K-nearest neighbor maupun metode lainnya. Studi literatur pada penelitian ini berupa teori-teori dari beberapa sumber seperti buku, jurnal dan referensi dari tugas tugas akhir mahasiswa yang berkenaan dengan klasifikasi varietas kopi.

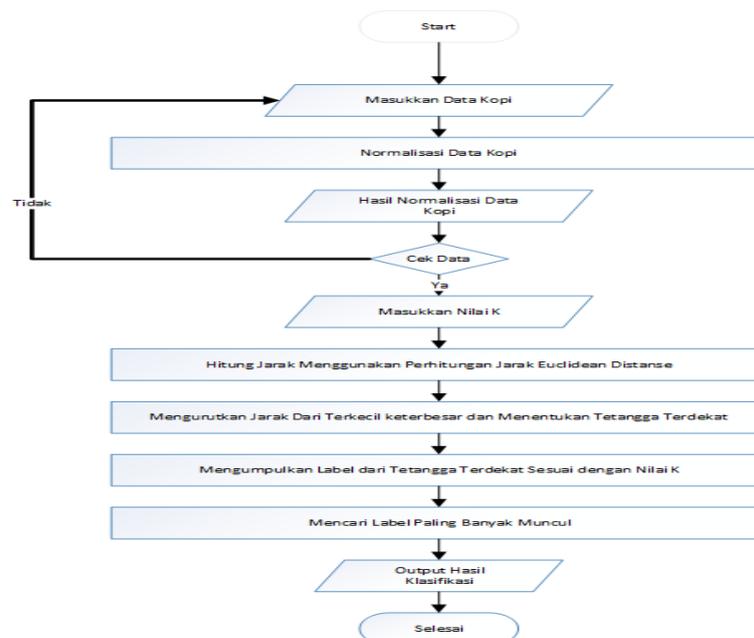
Tahap analisis kebutuhan merupakan tahap menganalisa kebutuhan yang akan dirancang dan dibangun. Setelah melakukan analisa maka langkah selanjutnya adalah merincika hasil analisa. Hasil analisa tersebut akan menjadi acuan dari perancangan sistem yang akan di bangun. Berikut hasil analisa kebutuhan sistem dalam penelitian tugas akhir ini.

Perangkat keras (*hardware*) merupakan perangkat yang sangat diperlukan dalam sistem komputer, pada penelitian tugas akhir ini perangkat keras yang digunakan dalam pembuatan dan perancangan sistem adalah Laptop Acer Aspire ES1-432 .

Perangkat lunak merupakan salah satu faktor pendukung untuk membuat dan merancang sebuah sistem. Perangkat lunak berfungsi sebagai pengolah data. Perancangan sistem dilakukan dengan beberapa tahap, yaitu perancangan arsitektur dan perancangan basis data. Perancangan arsitektur menghasilkan Diagram Konteks, perancangan basis data menghasilkan *Unified Modeling Language* (UML).

Pada tahap implementasi sistem rancangan yang telah dibuat pada tahap sebelumnya diimplementasikan dalam bahasa pemrograman. Output yang dihasilkan oleh sistem tersebut adalah "Unggul" dan jika tidak sesuai maka akan menghasilkan output "Kurang Unggul" hasil output ini tergantung kepada atribut yang ada pada sebuah data uji.

Pada tahap ini dilakukan pengujian sistem yang telah dilakukan pada tahap implementasi sistem. Tahap ini berguna untuk mengetahui apakah implementasi sistem yang dilakukan sudah sesuai dengan kebutuhan.



Gambar 1. Skema sistem

Analisa Dan Pembahasan

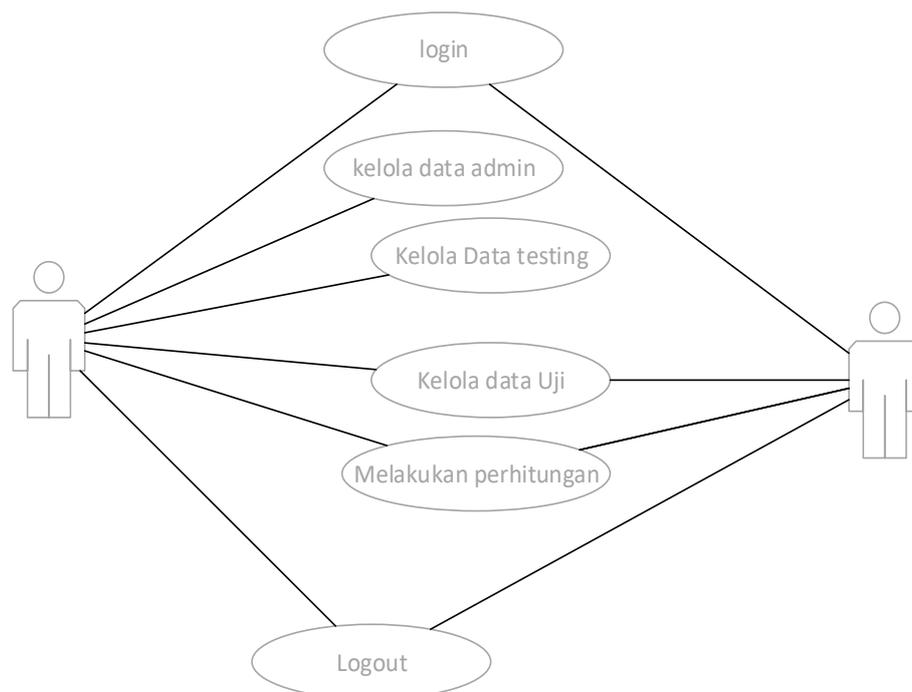
Analisis merupakan sebuah istilah yang secara kolektif mendeskripsikan fase- fase komponen dengan mempelajari seberapa bagus bagianbagian komponen tersebut awal pengembangan sistem. Analisis sistem adalah teknik pemecahan masalah bagian-bagian bekerja dan berinteraksi untuk mencapai tujuan. Sistem yang akan di bangun pada penelitian ini adalah sebuah sistem klasifikasi varietas kopi arabika menggunakan metode *K-Nearest Neighbour*. Sistem ini dibangun berbasis web dengan menggunakan bahasa pemrograman PHP dan database menggunakan *MYSQL*. Untuk mendapatkan hasil berupa klasifikasi kualitas varietas kopi arabika yang terbagi menjadi 3 kategori yaitu sangat unggul, unggul dan kurang unggul pengguna pada sistem ini terlebih dahulu melakukan pengisian data atribut-atribut yang berupa karakteristik morfologis kopi arabika.

Kabupaten Bener Meriah sebagai salah satu daerah penghasil kopi yang masyarakatnya sangat menggantungkan hidup dari hasil bertani kopi. Untuk memenuhi kebutuhan pasar yang terus meningkat baik pasar lokal maupun dunia diperlukan sebuah upaya untuk meningkatkan hasil atau produktivitas dari perkebunan kopi petani yang ada di kabupaten Bener Meriah khususnya. Salah satu cara yang dapat dilakukan salah satunya adalah dengan adanya bibit yang berkualitas atau varietas unggulan. Varietas yang dikembangkan oleh petani di kab. Bener Meriah sangat bervariasi, bahkan banyak dari varietas-varietas tersebut dicampur dalam satu lahan sempit yang sama. Akibatnya banyak dijumpai varietas-varietas kombinasi baru, yang mempunyai kualitas fisik, hasil produksi dan sistem klasifikasi untuk membedakan varietas kopi agar tidak mengurangi kualitas, cita rasa dan agar produksi kopi tetap terjaga kestabilannya.

Membangun suatu aplikasi berbasis web untuk mengklasifikasikan varietas kopi arabika. Dengan memasukkan data latih dan data testing kemudian melakukan perhitungan menggunakan metode *K-Nearest Neighbour* (KNN). Hasil perhitungan ditentukan dari kategori atau label mayoritas yang paling banyak muncul yang akan digunakan sebagai hasil klasifikasi.

Tahap selanjutnya adalah dengan melakukan perancangan sistem yang bertujuan sebagai penggambaran atau perencanaan untuk memenuhi kebutuhan sistem. Sistem ini di desain menggunakan perancangan *Unified Modeling Language* (UML) yang digunakan untuk menggambarkan mekanisme proses kerja dari sistem dan agar mempermudah dalam mengimplementasikan kedalam bahasa pemrograman.

Usecase adalah sebuah alat bantu guna menstimulasi pengguna untuk menggambarkan suatu sistem dari sudut pandangnya. Use case menggambarkan *external view* dari sistem yang akan kita buat modelnya (Muslihudin, 2016). Berikut ini adalah Use Case Diagram dari sistem Klasifikasi varietas kopi arabika menggunakan metode *K-Nearest Neighbour* (KNN)



Gambar 2. Use case Diagram

Tabel 1. Dataset

No	Varietas	Buah	Cabang	Bunga	Daun	Batang	Kategor
1	Catuai	15	25	13	9	170	Unggul
2	Caturra	14	25	14	11	197	Unggul
3	Cera	17	24	15	14	200	Sangat unggul
4	Costa Rica	14	13	13	13	200	Unggul
5	Dilla	16	22	14	15	170	Unggul
6	Ennarea	15	19	15	13	190	Sangat unggul
7	Ghimbi	14	10	14	9	300	Unggul
8	Harrar	16	29	11	13	197	Unggul
9	Jawa	16	15	10	13	170	Unggul
.....
207	Lekempti	14	17	18	6	250	Sangat unggul

Sebelum melakukan proses perhitungan dengan metode *K-nearest Neighbor*. Penulis membagi data menjadi 209 data testing dan 54 data uji dan melakukan proses normalisasi data terlebih dahulu yaitu dengan menggunakan metode normalisasi nilai maximal dan nilai minimal pada setiap kolom atribut. Adapun nilai maximal dan minimal pada tiap atribut didataset ditunjukkan oleh tabel berikut:

Tabel 2. Tabel Nilai Min Max

	Buah	Cabang	Bunga	Daun	Batang
max	25	30	18	15	300
min	12	6	6	5	150

Proses normalisasi data atribut buah, cabang, bunga, daun dan batang dengan nilai atribut 14, 9, 15, 7, 200 adalah sebagai berikut :

$$\text{Nilai atribut buah} = (14 - 12) / (25 - 12) = 0,153846154$$

$$\text{Nilai atribut cabang} = (9 - 6) / (30 - 6) = 0,125$$

$$\text{Nilai atribut bunga} = (15 - 6) / (18 - 6) = 0,75$$

$$\text{Nilai atribut daun} = (7 - 5) / (15 - 5) = 0,2$$

$$\text{Nilai atribut batang} = (200 - 150) / (300 - 150) = 0,333333$$

Perhitungan jarak euclidean distance antara data uji USDA dengan semua data testing yang memiliki nilai atribut buah 0,46153, cabang 0,46153, bunga 0,461538462, daun 0,4 dan batang 0,3333333.

Untuk klasifikasi menggunakan *K-Nearest Neighbour* perlu ditentukan banyaknya nilai K. Nilai K berfungsi untuk menentukan banyaknya nilai atau jarak terdekat dari hasil perhitungan.

Perhitungan jarak euclidean distance antara data uji 1 yaitu USDA dengan semua data testing yang memiliki nilai atribut buah 0,46153, cabang 0,5, bunga 0,416667, daun 0,4 dan batang 0,3333333 adalah sebagai berikut :

$$\text{dis} = \sqrt{(0,15384 - 0,46153) * 2 + (0,125 - 0,5) * 2 + (0,75 - 0,416667) * 2 + (0,2 - 0,4) * 2 + (0,333333333 - 0,333333333) * 2} = \mathbf{0,621619391}$$

Mengurutkan Hasil Perhitungan Dan Menentukan Tetangga Terdekat

Berdasarkan hasil dari perhitungan dengan perhitungan menggunakan rumus euclidean didapatkan tetangga terdekat dengan nilai k = 3 adalah sebagai berikut :

Tabel 3. Tabel Tetangga Terdekat K 3

Tetangga Terdekat	Jarak	Rangking
C47	0,152517554	1
Agaro	0,159471153	2
IAPAR 59	0,185068887	3

Menentukan Label Dari Tetangga Terdekat

Proses berikutnya dalam KNN adalah menentukan label atau kategori dari tetangga terdekat yang telah ditentukan ditunjukkan pada tabel dibawah ini :

Tabel 4. Tabel Label tetangga Terdekat

Tetangga Terdekat	Jarak	Rangking	Label
C47	0,152517554	1	Unggul
Agaro	0,159471153	2	Sangat Unggul
IAPAR 59	0,185068887	3	Sangat Unggul

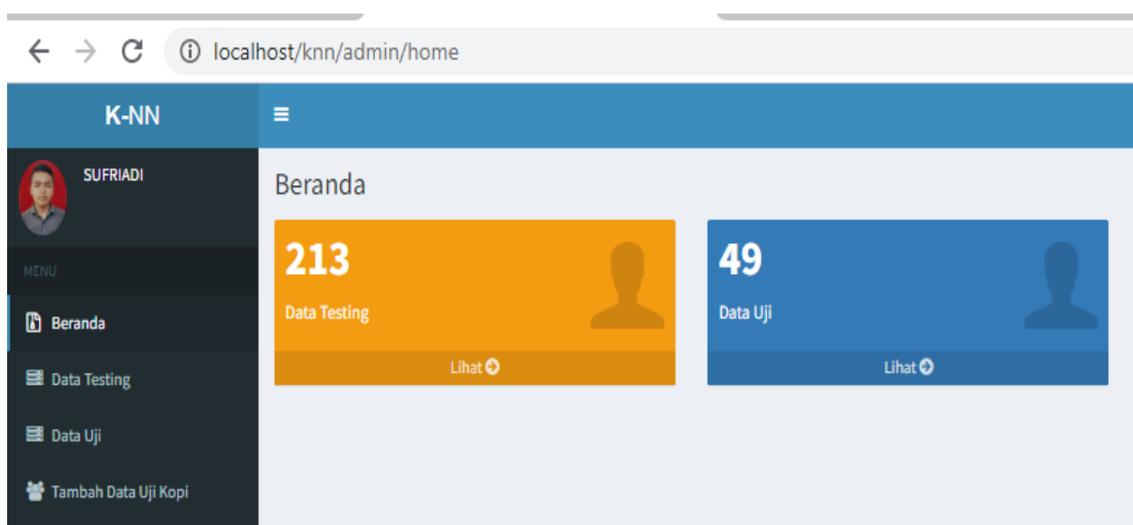
Gunakan Label Mayoritas sebagai hasil klasifikasi

Dari pengujian yang dilakukan dengan nilai k 3 terhadap data uji USDA dengan nilai atribut buah 0,461538462, cabang 0,5, bunga 0,416666667, daun 0,4 dan batang 0,333333333 label mayoritas yang paling banyak muncul dari nilai k=3 adalah sangat unggul. jadi varietas USDA diklasifikasikan sebagai sangat unggul.

Evaluasi sistem yaitu sebuah metode yang digunakan untuk mengukur kinerja sistem algoritma yang kita gunakan. Salah satu cara yang dapat dilakukan dalam mengukur kinerja sistem adalah dengan mengetahui tingkat akurasi dengan menggunakan metode confusion matriks.

$Accuracy = \frac{TP+TP+TP}{Total\ data}$

Halaman Beranda



Gambar 3. Halaman Beranda

Halaman Hasil Perhitungan

No	Nama	Jarak	Label
1	C47	0.1525	UNGGUL
2	Agaro	0.1594	SANGAT UNGGUL
3	IAPAR 59	0.1851	SANGAT UNGGUL

Gambar 4. Hasil Perhitungan

Kesimpulan

Berdasarkan hasil dari pembahasan dan pengujian sistem, maka didapatkan kesimpulan sebagai berikut:

1. Sistem klasifikasi varietas kopi arabika dengan menggunakan metode *K-Nearest Neighbour* memiliki atribut karakteristik morfologi varietas kopi arabika yaitu ukuran buah, cabang yang produktif, bunga, daun, dan batang yang diambil sebagai data karakteristik untuk dilakukan perhitungan.
2. Penelitian ini berhasil membangun sebuah sistem klasifikasi varietas arabika menggunakan metode *K-Neares Neighbour* yang digunakan untuk menentukan varietas kopi dengan memberikan hasil klasifikasi yaitu sangat unggul, unggul, dan kurang unggul.
3. Untuk kerja algoritma *K-Nearest Neighbour* yang diterapkan pada sistem klasifikasi varietas kopi arabika dari 54 data pengujian dengan nilai $k=3$ mendapatkan hasil 20 data sebagai kelas sangat unggul, 30 unggul dan 4 kurang unggul.
4. Tingkat keakurasian sistem klasifikasi varietas kopi arabika dengan menggunakan algoritma KNN dengan nilai $k=3$ didapatkan akurasi sebesar 70%.

DAFTAR PUSTAKA

- [1] Arifin, Z. W. J. (2019). *Penerapan Metode KNN (K-Nearest Neighbor) Dalam Sistem Pendukung Keputusan Penerima Kip (Kartu Indonesia Pintar) di Desa Pandean Berbasis Web Dan Mysql*. NJCA- volume 4, 27-34.
- [2] Hamni. (2013). *Potensi Pengembangan Teknologi Proses Produksi Kopi Lampung*. Jurnal Mechanical, Volume 4. Nomor 1.
- [3] Mustafa, R. &. (2017). *Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. Citec Journal, 153.
- [4] Novita, R. (2016). *Teknik Data Mining : Algoritma C4.5*. 11-12.
- [5] Wardani, S. H. (2018). *Klasifikasi Jenis Tanaman Mangrove Menggunakan Metode K-Nearest Neighbour*.
- [6] Widrawati, N. A. (2020). *Klasifikasi Varietas Kopi Arabika Menggunakan Metode Support Vektor Machine (SVM)*. Skripsi Sarjana. Universitas Sinartha Darma Yogyakarta.