

A Computer Vision Approach For Classifying California Papaya Ripeness Using K-Nearest Neighbor

Tyas Wulandari¹⁾, Iwan Ady Prabowo²⁾, Yustina Retno Wahyu Utami³⁾, Bayu Dwi Raharja⁴⁾, Hendro Wijayanto⁵⁾

^{1) 2) 3)} Program Studi Informatika, Universitas Tiga Serangkai Surakarta

^{4) 5)} Program Studi Teknologi Informasi, Universitas Tiga Serangkai Surakarta

¹⁾ wulandarityas.847@gmail.com; ²⁾ iwanadyp@tsu.ac.id; ³⁾ yustinaretno@tsu.ac.id; ⁴⁾ bayudr@tsu.ac.id;

⁵⁾ hendrow@tsu.ac.id

ABSTRACT

Determining the ripeness level of California papaya is important for harvest decisions, sorting, distribution, and spoilage control. In practice, ripeness identification is still commonly performed visually and is therefore subjective. This study aims to develop a digital image-based classification system for California papaya ripeness using the K-Nearest Neighbor (K-NN) algorithm with Hue and Saturation features in the Hue Saturation Value (HSV) color space. The dataset consists of 90 primary images, divided into 60 training images and 30 testing images, with four ripeness classes: unripe, half-ripe, ripe, and rotten. All images were captured using a Xiaomi Mi A2 Lite smartphone and cropped to 1436 × 1000 pixels. Classification was conducted using Euclidean distance. The value of k was selected empirically through trial and error in the original study, and k = 9 was retained because it produced the most stable result on the available data while reducing the potential for class ties. The evaluation produced 22 correct predictions out of 30 test images, resulting in an accuracy of 73.33%. This revised manuscript strengthens the methodological reporting by clarifying parameter selection, documenting the data distribution and providing a literature-based comparison with alternative methods, such as Support Vector Machine (SVM) and Convolutional Neural Network (CNN). The findings suggest that K-NN with HSV features remains a feasible, low-cost baseline, although its performance should be improved through larger datasets, per-class evaluation reporting, and head-to-head comparisons on the same dataset.

Keywords: California papaya, ripeness classification, image processing, HSV, K-Nearest Neighbor

I. PENDAHULUAN

Pepaya merupakan komoditas hortikultura tropis yang bernilai ekonomi tinggi dan banyak dibudidayakan di Indonesia. Buah pepaya merupakan salah satu hasil pertanian terbesar di daerah Mojosongo Boyolali, yang biasa dikonsumsi oleh masyarakat. Permintaan pasar untuk komoditas pepaya masih tinggi dan belum mampu memenuhi kebutuhan konsumen[1]. Tingginya permintaan pasar dan distribusi pepaya mengharuskan petani mampu membedakan tingkat kematangan buah pepaya untuk mengurangi risiko pembusukan.

Dalam rantai pasok buah segar, tingkat kematangan berpengaruh langsung terhadap kualitas visual, rasa, tekstur, masa simpan, dan ketepatan distribusi. Pada komoditas pepaya, warna kulit buah menjadi indikator visual utama untuk membedakan kondisi kematangan buah [2]. Namun, penilaian manual oleh petani atau pedagang sangat dipengaruhi oleh pengalaman, pencahayaan, kelelahan, dan

persepsi individu, sehingga hasil klasifikasi sering kali tidak konsisten.

Dari permasalahan tersebut, penelitian ini mengusulkan metode untuk mengklasifikasikan tingkat kematangan buah pepaya California. Penelitian dengan memanfaatkan fitur warna *Red Green Blue* (RGB) telah banyak dilakukan [2], [3]. Untuk memperkaya metodologi, penelitian ini memanfaatkan fitur warna HSV. Algoritma *K-Nearest Neighbor* dipilih karena algoritma ini dapat digunakan untuk mengolah data yang bersifat numerik dan tidak membutuhkan skema estimasi parameter perulangan yang rumit.

II. TINJAUAN PUSTAKA

Kajian mutakhir menunjukkan bahwa klasifikasi kematangan buah telah berkembang pesat melalui kombinasi pengolahan citra, *machine learning*, *deep learning*, dan sensor nondestruktif. Fitur warna masih relevan untuk banyak buah dengan perubahan visual yang jelas selama pematangan[4], sedangkan review

terbaru menekankan pentingnya standardisasi dataset dan metrik evaluasi agar hasil antar penelitian dapat dibandingkan secara lebih adil [5]. Di tingkat aplikasi, pengujian akurasi, precision, recall, dan F1-score menjadi metrik yang paling umum digunakan untuk evaluasi klasifikasi kematangan buah dan sayuran [6].

Pada objek pepaya, sejumlah penelitian telah menggunakan pendekatan berbeda. Pemanfaatan ruang warna HSV dan Naive Bayes untuk klasifikasi kematangan papaya menghasilkan akurasi yang tinggi [7] dan diperluas dengan fitur warna, tekstur, dan bentuk menggunakan SVM [3]. Pendekatan CNN menghasilkan akurasi tinggi pada klasifikasi kematangan pepaya [8], [9]. Sementara itu, dilaporkan akurasi pada klasifikasi kematangan pepaya California berbasis fitur RGB-HSV menggunakan SVM sebesar 80% [10]. Temuan-temuan tersebut memperlihatkan bahwa model yang lebih kompleks cenderung memberikan performa lebih tinggi, tetapi memerlukan data, komputasi, dan prosedur pelatihan yang lebih besar [11], [12], [13].

Di sisi lain, K-NN tetap penting sebagai baseline karena sederhana, mudah diimplementasikan, dan sesuai untuk dataset kecil. Kelemahannya terletak pada sensitivitas terhadap pemilihan nilai k, ketidakseimbangan kelas, dan overlap fitur antar kelas. Oleh sebab itu, penelitian ini bertujuan memperkuat kualitas ilmiah penelitian melalui perbaikan terminologi, penguatan justifikasi pemilihan $k = 9$, penambahan pelaporan evaluasi klasifikasi, penyajian distribusi data yang terdokumentasi, serta perbandingan hasil dengan sedikitnya dua pendekatan lain berbasis literatur.

III. METODE PENELITIAN

3.1 Jenis Data dan Akuisisi Citra

Penelitian ini menggunakan data primer berupa citra buah pepaya California yang diperoleh melalui pengambilan gambar langsung. Akuisisi data lapangan diperkuat oleh wawancara dengan petani sekaligus distributor pepaya California di Butuh, Mojosongo, Boyolali. Seluruh citra diambil menggunakan smartphone Xiaomi Mi A2 Lite, disimpan dalam format JPG, kemudian dipotong (*cropping*) hingga ukuran 1436×1000 piksel untuk menyesuaikan area analisis.

Dataset terdiri atas 90 citra yang dibagi menjadi 60 data latih dan 30 data uji. Penelitian membedakan empat kelas kematangan, yaitu mentah, setengah matang, matang, dan busuk. Variabel yang digunakan adalah rata-rata nilai

Hue dan rata-rata nilai Saturation dari setiap citra. Dengan dua fitur tersebut, model berupaya mengenali perubahan dominan warna kulit pepaya sepanjang proses pematangan. Profil dataset diperlihatkan pada Tabel 1.

Tabel 1. Profil dataset dan distribusi data yang terdokumentasi

Komponen	Keterangan	Nilai
Jumlah citra	Total dataset penelitian	90 citra
Data latih	Digunakan untuk basis referensi K-NN	60 citra
Data uji	Digunakan untuk evaluasi model	30 citra
Jumlah kelas	Kategori tingkat kematangan	4 kelas
Label kelas	Kelas yang digunakan dalam penelitian	Mentah, setengah matang, matang, busuk
Ukuran citra	Hasil setelah cropping	1436×1000 piksel
Fitur	Variabel numerik untuk klasifikasi	Rata-rata Hue dan rata-rata Saturation
Distribusi per kelas pada data latih	Status dokumentasi arsip penelitian awal	Tidak terdokumentasi rinci
Distribusi per kelas pada data uji	Status dokumentasi arsip penelitian awal	Tidak terdokumentasi rinci

3.2 Prapemrosesan dan Ekstraksi Fitur

Tahap prapemrosesan dilakukan melalui segmentasi citra untuk mempertahankan area visual yang relevan. Segmentasi dan ekstraksi fitur menggunakan Matlab 2019. Penelitian awal juga menyebutkan penggunaan ambang rata-rata pada ruang warna RGB untuk mendukung proses segmentasi. Setelah itu, citra dikonversi ke ruang warna HSV untuk memperoleh dua fitur inti, yakni rata-rata *Hue* dan rata-rata *Saturation*. Pemilihan HSV didasarkan pada kenyataan bahwa perubahan warna kulit pepaya selama pematangan lebih mudah direpresentasikan secara perseptual dalam ruang warna ini dibandingkan dengan hanya menggunakan RGB.

3.3 Klasifikasi K-Nearest Neighbor

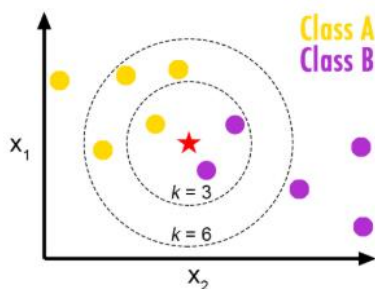
K-Nearest Neighbor, yang sering disingkat k-NN, adalah sebuah teknik dalam pembelajaran mesin yang memanfaatkan jarak atau kedekatan antar data untuk menentukan kategori atau nilai dari data baru. Dalam metode ini, sebuah titik data akan dianalisis dengan melihat k tetangga terdekatnya, kemudian

keputusan klasifikasi diambil berdasarkan mayoritas kategori tetangga tersebut atau rata-rata nilai mereka. Metode ini bersifat nonparametrik, artinya tidak memerlukan asumsi awal tentang distribusi data. Kelebihan k-NN termasuk kesederhanaannya dan kemampuan menangani masalah multi-kelas [14].

K-NN mengklasifikasikan data uji berdasarkan mayoritas label dari k tetangga terdekat dalam ruang fitur. Jarak antar data dihitung menggunakan *Euclidean distance* (rumus 1).

$$euc = \sqrt{(a_1 - b_1)^2 + \dots + (a_n - b_n)^2} \quad (1)$$

Sebagai ilustrasi, setiap citra direpresentasikan oleh dua fitur numerik, sehingga jarak antar citra didefinisikan pada ruang dua dimensi seperti diperlihatkan pada Gambar 1. Setelah seluruh jarak dihitung, data latih diurutkan dari jarak terkecil hingga terbesar, lalu kelas mayoritas dari k tetangga terdekat ditetapkan sebagai hasil klasifikasi citra uji.



Gambar 1. Penggunaan k pada metode K-NN [14]

Contoh perhitungan K-NN mengambil salah satu data uji dengan nilai Hue 0,1855 dan Saturation 0,5273. Setelah dihitung terhadap seluruh data latih, sembilan tetangga terdekat menunjukkan dominasi kelas setengah matang sehingga citra tersebut diklasifikasikan sebagai setengah matang.

3.4 Justifikasi Pemilihan Nilai k = 9

Nilai k dipilih melalui trial-and-error, alasan pemilihan k diperkuat agar lebih metodologis. Pertama, penggunaan nilai ganjil mengurangi kemungkinan terjadinya tie pada proses voting mayoritas. Kedua, dengan jumlah data latih 60 citra dan dua fitur numerik, nilai k yang terlalu kecil akan membuat keputusan sangat sensitif terhadap noise lokal, sedangkan nilai k yang terlalu besar berpotensi mengaburkan batas antar kelas yang memang berdekatan secara visual, terutama antara kelas mentah, setengah matang, dan matang. Ketiga, k = 9 berada pada

tingkat moderat, cukup lokal untuk mempertahankan karakteristik data terdekat, tetapi cukup stabil untuk meredam pengaruh 1 atau 2 outlier. Oleh karena itu, nilai k = 9 dipertahankan sebagai parameter final karena paling sesuai dengan hasil trial-and-error pada penelitian awal. Untuk memilih k seharusnya menggunakan validation set atau k-fold *cross-validation*. Dengan demikian, pada penelitian lanjutan, pemilihan k tidak hanya berdasarkan trial-and-error pada satu pembagian data, tetapi juga berdasarkan performa yang lebih robust pada beberapa lipatan data.

IV. HASIL DAN PEMBAHASAN

4.1 Perhitungan Klasifikasi K-Nearest Neighbor

Klasifikasi kematangan pepaya menggunakan 60 data training yang diperlihatkan pada Tabel 2.

Tabel 2. Data Training

NO	INPUT		OUTPUT
	HUE	SATURATION	KEMATANGAN
1	0.2783	0.3744	MENTAH
2	0.1659	0.4493	ST MATANG
3	0.1872	0.566	ST MATANG
4	0.2373	0.4775	MENTAH
5	0.1286	0.6848	MATANG
6	0.1223	0.5814	MATANG
7	0.1323	0.4687	BUSUK
8	0.2335	0.3452	BUSUK
9	0.2381	0.5024	MENTAH
10	0.175	0.6588	ST MATANG
11	0.2322	0.4773	MENTAH
...
...
30	0.1165	0.6664	MATANG
31	0.1341	0.72	MATANG
32	0.1503	0.4357	BUSUK
...
...
60	0.0919	0.6867	BUSUK

Jika terdapat data ke 61 dengan nilai rata-rata hue 0.1855 dan saturation 0.5273, maka jarak data ke 61 terhadap data training menggunakan rumus (1). Contoh perhitungan jarak Euclidean data-61 ke data-1 sebagai berikut:

$$d_1, d_{61} = \sqrt{(0,2783 - 0,1855)^2 + (0,3744 - 0,5273)^2} = 0,178858184$$

Urutan jarak terdekat data-61 terhadap data training dengan $k=9$ diperlihatkan pada Tabel 3.

Tabel 3. Jarak terdekat $k=9$

Data ke-	Jarak euclidean	Kategori
59	0.033060551	BUSUK
45	0.033370945	ST MATANG
13	0.0344308	ST MATANG
3	0.038737321	ST MATANG
56	0.047089383	ST MATANG
21	0.048423238	BUSUK
28	0.048658504	ST MATANG
9	0.058195962	MENTAH
12	0.064301866	MATANG

Berdasarkan jarak terdekat pada Tabel 4, data 61 dapat diklasifikasikan ke dalam kategori setengah matang.

4.2 Evaluasi Klasifikasi

Evaluasi klasifikasi menggunakan data uji sebanyak 30 citra. Performa model dalam penelitian ini diperoleh nilai 73,33% dengan 22 prediksi yang benar dari 30 data uji. Selain itu, tingkat kesalahan penelitian ini sebesar 26,67%.

4.3 Interpretasi Hasil Klasifikasi

Nilai akurasi 73,33% menunjukkan bahwa pendekatan K-NN berbasis fitur rata-rata *hue* dan *saturation* mampu menangkap pola umum perubahan warna kulit pepaya California, tetapi performanya masih berada pada tingkat moderat. Hasil ini cukup layak untuk baseline sistem atau prototipe bantuan sortasi awal, namun belum cukup kuat untuk menjadi satu-satunya dasar keputusan pada skenario implementasi yang menuntut reliabilitas tinggi. Overlap visual antarkelas, terutama antara mentah, setengah matang, dan matang, diduga menjadi penyebab utama kesalahan klasifikasi.

Secara khusus, kelas setengah matang merupakan fase transisi yang secara visual sering berbagi karakteristik dengan kelas mentah maupun matang. Dalam kondisi seperti ini, dua fitur warna sederhana belum selalu memadai untuk membentuk batas keputusan yang tegas. Kelas busuk juga berpotensi tumpang tindih dengan kelas matang bila perubahan warna belum terlalu ekstrem atau kondisi pencahayaan tidak seragam saat pengambilan citra. Dengan demikian, hasil penelitian memperlihatkan bahwa representasi warna saja dapat bekerja sebagai baseline,

tetapi masih perlu dilengkapi dengan fitur tekstur, bentuk, atau pendekatan model yang lebih ekspresif.

4.4 Kekuatan dan Keterbatasan Nilai $k = 9$

Pemilihan $k = 9$ dapat dibenarkan secara praktis karena menghasilkan klasifikasi mayoritas yang relatif stabil pada contoh perhitungan dan pada evaluasi secara keseluruhan. Nilai ini juga mengurangi dominasi satu tetangga ekstrem yang mungkin muncul pada k kecil. Namun, karena data latih hanya berjumlah 60 citra dan distribusi per kelas tidak terdokumentasi secara rinci, $k = 9$ tetap perlu dipandang sebagai hasil tuning empiris yang bergantung pada dataset. Artinya, pada dataset lain atau kombinasi fitur yang berbeda, nilai k optimal dapat berubah.

4.5 Perbandingan dengan Algoritma Lain

Untuk memperkuat posisi temuan, penelitian ini dibandingkan secara literatur dengan pendekatan lain pada tugas klasifikasi kematangan pepaya atau tugas serupa. Perlu ditegaskan bahwa perbandingan pada Tabel 4 bukan *head-to-head* karena dataset, jumlah kelas, fitur, dan skenario eksperimen tidak identik. Meski demikian, tabel tersebut tetap bermanfaat untuk menunjukkan posisi performa model K-NN sederhana yang digunakan dalam penelitian ini.

Tabel 4. Perbandingan berbasis literatur dengan algoritma lain

Penelitian	Metode	Objek/ Fitur	Hasil utama	Catatan
Penelitian ini	K-NN	Pepaya California; mean Hue dan Saturation	Akurasi 73,33%	4 kelas; 90 citra; baseline sederhana
Wardani dkk. [3]	SVM	Pepaya; warna, tekstur, bentuk	Model lebih kaya fitur	Menunjukkan arah penguatan melalui multi-fitur
Maito dkk. [10]	SVM	Pepaya California; RGB dan HSV	Akurasi 80%	150 citra; 3 kelas; 80% train, 20% test
Nurmalasari dkk. [8]	CNN	Pepaya; warna dan tekstur	Akurasi 97%	Transfer learning dan simulasi yang lebih intensif
Hawibowo dan Muhimmah [9]	CNN	Pepaya; citra untuk Android	Akurasi 96,97%	Implementasi aplikasi Android; 3 kelas

Tabel 4 memperlihatkan bahwa K-NN pada penelitian ini masih tertinggal dibandingkan dengan SVM dan CNN yang dilaporkan dalam studi lain. Hal tersebut dapat dipahami karena penelitian ini hanya menggunakan dua fitur warna rata-rata, ukuran dataset relatif kecil, serta tidak didukung oleh augmentasi data atau ekstraksi fitur yang lebih kaya. Namun demikian, keunggulan K-NN terletak pada kesederhanaan implementasi, transparansi proses pengambilan keputusan, dan kebutuhan komputasi yang rendah. Dalam konteks laboratorium dengan sumber daya terbatas, model ini tetap relevan sebagai baseline atau titik awal sebelum beralih ke model yang lebih kompleks.

V. PENUTUP

Penelitian ini berhasil membangun sistem klasifikasi tingkat kematangan buah pepaya California menggunakan algoritma K-Nearest Neighbor berbasis fitur warna Hue dan Saturation pada ruang warna HSV. Dataset penelitian terdiri atas 90 citra, yaitu 60 data latih dan 30 data uji, dengan empat kelas kematangan: mentah, setengah matang, matang, dan busuk. Hasil evaluasi menunjukkan 22 prediksi benar dari 30 data uji, sehingga diperoleh akurasi 73,33%.

Pemilihan nilai $k = 9$ dipertahankan karena secara empiris memberikan hasil yang paling stabil pada penelitian awal dan secara metodologis masih masuk akal untuk ukuran dataset yang digunakan. Hasil penelitian menegaskan bahwa K-NN dengan fitur warna sederhana tetap layak sebagai *baseline* yang murah dan mudah diterapkan. Namun, performanya masih berada di bawah pendekatan yang lebih kaya fitur atau lebih kompleks seperti SVM dan CNN dalam studi lain. Oleh sebab itu, penelitian lanjutan disarankan menambah dataset, mencatat distribusi kelas secara rinci, melaporkan confusion matrix per kelas, serta melakukan perbandingan *head-to-head* antar algoritma pada dataset yang sama.

DAFTAR PUSTAKA

[1] H. Miftah *et al.*, "Implementasi Rantai Pasok Pepaya California Yang Berpihak Pada Petani Gabungan Kelompok Tani (Gapoktan)," *Jurnal Qardhul Hasan*, vol. 8, no. 1, pp. 20–25, Apr. 2022, doi: <http://dx.doi.org/10.20961/agritexts.v4i6i2.67123>.

- [2] K. Utama Putra, W. Yosfand, and A. Ramadhani, "Klasifikasi Kematangan Buah Pepaya Berdasarkan Warna Menggunakan Convolutional Neural Network," *JITSI: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 6, no. 1, pp. 1–6, Mar. 2025, doi: 10.62527/jitsi.6.1.283.
- [3] L. A. Wardani, G. Pasek, S. Wijaya, and F. Bimantoro, "Klasifikasi Jenis Dan Tingkat Kematangan Buah Pepaya Berdasarkan Fitur Warna, Tekstur Dan Bentuk Menggunakan Support Vector Machine," *Jurnal Teknologi Informasi, Komputer dan Aplikasinya*, vol. 4, no. 1, pp. 75–85, Mar. 2022, doi: 10.29303/JTIKA.V4I1.171.
- [4] M. Rizzo, M. Marcuzzo, A. Zangari, A. Gasparetto, and A. Albarelli, "Fruit ripeness classification: A survey," Mar. 01, 2023, *KeAi Communications Co.* doi: 10.1016/j.aiaa.2023.02.004.
- [5] Z. F. R. Ahmed, A. K. Abdalla, N. Kaur, and F. Wu, "Insights into recent developments and obstacles in automated fruit ripeness classification," Apr. 01, 2026, *KeAi Communications Co.* doi: 10.1016/j.grets.2025.100302.
- [6] R. Sakriawindarta and K. Kusriani, "Metode Klasifikasi Tingkat Kematangan Buah dan Sayuran: Tinjauan Sistematis," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 4, pp. 2344–2354, Oct. 2024, doi: 10.70609/gtech.v8i4.5067.
- [7] Ellif, S. H. Sitorus, and R. Hidayati, "Klasifikasi Kematangan Pepaya Menggunakan Ruang Warna HSV Dan Metode Naive Bayes Classifier," *Jurnal Komputer dan Aplikasi*, vol. 09, no. 01, pp. 66–75, 2021, doi: <https://doi.org/10.26418/coding.v9i01.45906>.
- [8] N. Nurmalasari, Y. A. Setiawan, W. Astuti, M. R. R. Saelan, S. Masturoh, and T. Haryanti, "Classification for Papaya Fruit Maturity Level With Convolutional Neural Network," *Jurnal Riset Informatika*, vol. 5, no. 3, pp. 331–338, Jun. 2023, doi: 10.34288/jri.v5i3.225.
- [9] M. Sayyidin, H. #1, and I. Muhimmah, "Aplikasi Pendeteksi Tingkat Kematangan Pepaya menggunakan Metode Convolutional Neural Network (CNN) Berbasis Android," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 10, no. 1, pp. 162–170, Apr. 2024.
- [10] R. M. Maito, M. Qamal, and Fajriana, "Identification of Papaya Ripeness Using the Support Vector Machine Algorithm,"

International Journal of Engineering, Science and Information Technology, vol. 5, no. 1, pp. 272–277, 2025, doi: 10.52088/ijesty.v5i1.710.

- [11] N. Aherwadi, U. Mittal, J. Singla, N. Z. Jhanjhi, A. Yassine, and M. S. Hossain, "Prediction of Fruit Maturity, Quality, and Its Life Using Deep Learning Algorithms," *Electronics (Switzerland)*, vol. 11, no. 24, Dec. 2022, doi: 10.3390/electronics11244100.
- [12] N. Ismail and O. A. Malik, "Real-time visual inspection system for grading fruits using computer vision and deep learning techniques," *Information Processing in Agriculture*, vol. 9, no. 1, pp. 24–37, Mar. 2022, doi: 10.1016/j.inpa.2021.01.005.
- [13] F. Xiao, H. Wang, Y. Li, Y. Cao, X. Lv, and G. Xu, "Object Detection and Recognition Techniques Based on Digital Image Processing and Traditional Machine Learning for Fruit and Vegetable Harvesting Robots: An Overview and Review," *Agronomy*, vol. 13, no. 3, pp. 1–29, Mar. 2023, doi: 10.3390/agronomy13030639.
- [14] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *ILKOM Jurnal Ilmiah*, vol. 11, no. 3, pp. 269–274, Dec. 2019, doi: 10.33096/ilkom.v11i3.489.269-274.