

Klasifikasi Bumbu Dapur Indonesia Menggunakan Metode *K-Nearest Neighbors* (K-NN)

Suastika Yulia Riska¹, Lia Farokhah²

^{1,2}Institut Teknologi dan Bisnis Asia Malang
¹riska.suastika@asia.com, ²farokhah@asia.ac.id

ABSTRAK

Bumbu adalah salah satu elemen yang sangat penting dalam sebuah masakan. Bumbu dapur atau rempah Indonesia memiliki jenis ragam yang sangat banyak. Kesalahan dalam pemilihan bumbu sangat berpengaruh terhadap rasa masakan. Image processing merupakan salah satu cabang ilmu dalam bidang teknologi yang dapat dimanfaatkan untuk mengenali objek citra yang ditangkap oleh kamera. Penelitian ini akan mengklasifikasikan jenis bumbu dapur yang hampir mirip yaitu jahe, lengkuas, kunyit dan kencur. Metode klasifikasi yang dipakai adalah *K-Nearest Neighbors* (K-NN). Pada penelitian ini kita menguji cara *split data training* dan *data testing* yaitu 66,7%: 33,33%, 75%:25% dan 90%:10%. Pembagian *data training* dan *data testing* menggunakan 90%:10% memiliki rata rata akurasi yang paling besar dibandingkan cara pembagian yang lain. Pemilihan K=3 atau K=5 memiliki rata rata akurasi yang hampir sama pada semua cara *split data training* dan *data testing* yaitu 64,66%: 65%. Pada K=1 memiliki akurasi yang cukup tinggi dibandingkan K sebelumnya yaitu 73%.

Kata Kunci: pembagian dataset, *K-Nearest Neighbors*, bumbu dapur Indonesia, Image Processing

ABSTRACT

Seasoning is one of the most important elements in a dish. Indonesian herbs or spices have a very wide variety of types. Mistakes in choosing spices have a big effect on the taste of the dish. Image processing is a branch of science in the field of technology that can be used to recognize image objects captured by the camera. This study will classify the types of spices that are almost similar, namely ginger, galangal, turmeric and kencur. The classification method used is *K-Nearest Neighbor* (K-NN). In this study we tested how to split training data and data testing, namely 66.7%: 33.33%, 75%: 25% and 90%: 10%. The sharing of training data and testing data uses 90%: 10% has the greatest average accuracy compared to other distribution methods. The selection of $K = 3$ or $K = 5$ has an average accuracy that is almost the same in all methods of split training data and testing data, namely 64.66%: 65%. At $K = 1$ it has a fairly high accuracy compared to the previous K , which is 73%.

Keywords: Split dataset, *K-Nearest Neighbors*, Indonesian kitchen seasoning, Image Processing

1. PENDAHULUAN

Salah satu hasil alam di Indonesia yang melimpah adalah rempah-rempah. Rempah-rempah yang dimiliki oleh Indonesia juga sangat beragam. Adapun jenis rempah-rempah yang ada di Indonesia antara lain: jahe, kunyit, lengkuas, kencur, dan masih banyak lagi. Adanya keberagaman rempah-rempah juga memiliki banyak manfaat untuk masyarakat Indonesia. Rempah-rempah tersebut banyak digunakan untuk penyedap masakan. Seiring dengan perkembangan ilmu pengetahuan, membuat masyarakat juga memiliki kesadaran tentang pentingnya kesehatan. Selain digunakan untuk penyedap masakan, rempah-rempah juga bisa digunakan untuk obat tradisional yang bermanfaat untuk kesehatan. Tidak hanya sebagai obat, rempah-rempah juga digunakan bermanfaat untuk menjaga imunitas. Namun dengan keberagaman tersebut juga berdampak kepada masyarakat modern. Banyak masyarakat yang masih merasa bingung membedakan setiap nama rempah-rempah. Di era yang semakin modern ini, semakin banyak teknologi yang berkembang, sehingga dapat dengan mudah dimanfaatkan oleh masyarakat. Adanya kesulitan

yang dialami masyarakat untuk membedakan rempah-rempah, diperlukan teknologi yang dapat dengan mudah mengenali nama setiap rempah-rempah.

Salah satu teknologi yang bisa dikembangkan adalah teknologi dengan menerapkan *image processing*. *Image processing* merupakan salah satu cabang ilmu dalam bidang teknologi yang dapat dimanfaatkan untuk mengenali objek citra yang ditangkap oleh kamera. Beragamnya metode untuk mengenali objek citra yang terdapat dalam image processing, membuat teknologi ini semakin banyak digunakan oleh peneliti. Cara pengenalan objek pada *image processing* disebut dengan metode klasifikasi, dimana dengan menggunakan metode klasifikasi dapat dilakukan pengelompokan dalam kelas-kelas tertentu [1]. Untuk mendapatkan hasil akurasi yang baik dalam penelitian juga perlu dilakukan analisis terhadap dataset penelitian. Terdapat 3 ekstraksi fitur dalam pengolahan citra digital, yaitu ekstraksi fitur warna, bentuk, dan tekstur. Dengan melakukan analisis terhadap ciri dataset yang digunakan, maka bisa memaksimalkan nilai akurasi.

Banyak metode klasifikasi yang telah digunakan, salah satunya adalah metode *K-Nearest Neighbors*(K-NN). K-NN memiliki beberapa keunggulan yaitu mudah untuk diimplementasikan, memiliki hasil yang baik untuk dataset yang memiliki banyak *noise*, dan sesuai untuk dataset dengan jumlah data *training* yang besar[2]. Metode ini juga sudah digunakan untuk proses klasifikasi di berbagai obyek penelitian. Dengan menggunakan metode K-NN penggunaan dataset penelitian juga sangat berpengaruh untuk menghasilkan nilai akurasi yang tinggi [1][3]. Widiarti [3] mengungkapkan bahwa metode K-NN merupakan metode yang sederhana dalam proses klasifikasi. Penelitian yang dilakukan tentang alihaksara citra aksara dari Batak, Jawa, dan Sunda mendapatkan hasil klasifikasi yang baik. Penelitian dengan menggunakan metode klasifikasi K-NN juga telah dilakukan untuk mengelompokkan jenis tumbuhan Mangrove. Hasil yang diperoleh dari hasil penelitian ini juga cukup baik dengan memperhitungkan beberapa karakteristik morfologi [4]. Klasifikasi dengan K-NN juga diterapkan untuk membedakan jenis daging sapi dan babi [5]. Pada penelitian tersebut digunakan ekstraksi fitur warna dengan HSV. Selain itu juga menggunakan ekstraksi fitur tekstur. Hasil akurasi untuk klasifikasi menunjukkan hasil 88,75%. Penggunaan ekstraksi fitur yang tepat juga sangat mempengaruhi hasil dari penelitian. Penelitian tentang pencarian resep berdasarkan ekstraksi fitur warna dan bentuk menghasilkan nilai MAP yang tinggi yaitu 94,1892% [6]. Pada penelitian ini menggunakan penggabungan dua ekstraksi fitur karena dataset memiliki karakteristik warna dan bentuk yang hampir sama. Klasifikasi dengan metode K-NN juga diterapkan untuk mengelompokkan zat penghambat pada protein kinase[7]. Pada penelitian tersebut, melakukan perbandingan dari beberapa metode klasifikasi, yaitu SVM dan Naïve Bayesian. Hasil akurasi menunjukkan metode klasifikasi dengan K-NN yang tertinggi yaitu 90% pada nilai K=7.

Penelitian-penelitian sebelumnya menunjukkan hasil yang baik dengan menerapkan metode klasifikasi KNN. Sulitnya membedakan rempah-rempah masih menjadi permasalahan yang krusial pada masyarakat modern. Sehingga, adanya permasalahan tentang sulitnya masyarakat modern mengenali rempah-rempah, perlu dilakukan penelitian berbasis image processing dengan menggunakan metode klasifikasi K-NN.

2. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini adalah *image* berbagai macam rempah-rempah sebanyak 800 citra. Adapun jenis rempah-rempah yang digunakan pada penelitian ini antara lain: jahe, kencur, kunyit, dan lengkuas. Setiap jenis rempah-

rempah terdiri dari 200 data citra. Dataset yang digunakan dalam penelitian ini diambil dari google. Teknik pengambilan dataset adalah dengan Teknik augmentasi. Dataset citra jahe, kencur, kunyit, dan lengkuas berturut-turut ditunjukkan pada Gambar 1, 2, 3, dan 4.



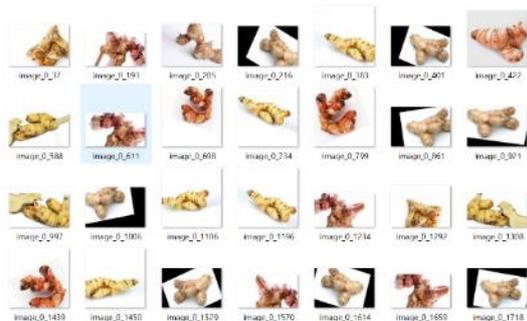
Gambar 1. Dataset Citra Jahe



Gambar 2. Dataset Citra Kencur

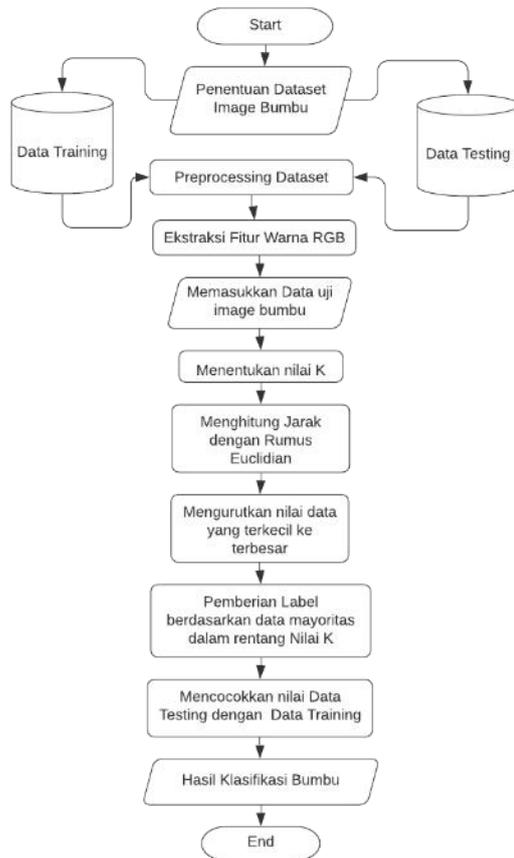


Gambar 3. Dataset Citra Kunyit



Gambar 4. Dataset Citra Lengkuas

Alur sistem penelitian ini ditunjukkan pada Gambar 5.



Gambar 5. Alur Sistem Penelitian

Berdasarkan Gambar 5 saat penentuan dataset dilakukan pembagian data *training* dan data *testing*. Pembagian data *training* dan data *testing* yang paling umum dibagi menjadi 3[8]. Pembagian ini digambarkan pada Tabel 1.

Tabel 1. Pembagian Porsi Data *Training* dan *Testing*

No	Data Training	Data Testing
1	66,7%	33,33%
2	75%	25%
3	90%	10%

Tahap *preprocessing* dilakukan untuk peningkatan kualitas dari citra bumbu, agar bisa mendapatkan nilai akurasi yang baik. Tahap selanjutnya adalah ekstraksi fitur warna. Ekstraksi fitur warna yang digunakan di dalam penelitian ini adalah ekstraksi fitur warna RGB. Ekstraksi ciri warna dilakukan dengan menggunakan normalisasi RGB. Nilai persentase *Red* ditunjukkan pada Persamaan (1). Nilai persentase *Green* ditunjukkan pada Persamaan (2). Nilai persentase *Blue* ditunjukkan pada Persamaan (3).

$$R = \frac{r}{r+g+b} \quad (1)$$

$$G = \frac{g}{r+g+b} \quad (2)$$

$$B = \frac{b}{r+g+b} \quad (3)$$

Dengan *r* adalah nilai *Red*, *g* adalah nilai *Green*, dan *b* adalah nilai *Blue*. Masing-masing komponen RGB menggunakan 8 bit, yaitu mencapai 255 untuk masing-masing komponen [6]. Pada penelitian ini tahap klasifikasi menggunakan metode KNN, dengan nilai *K*=1, *K*=3, dan *K*=5. Metode KNN merupakan salah satu metode klasifikasi dengan menghitung jarak ketetanggaan terdekat dari obyek. Untuk menentukan nilai *k* pada suatu penelitian adalah disesuaikan dengan dataset yang digunakan pada penelitian. nilai [9].

KNN melakukan klasifikasi terhadap citra uji pada kelas-kelas dengan jumlah anggota terbanyak. Nilai ketetanggaan dihitung dengan menggunakan Euclidean distance yang ditunjukkan pada Persamaan 4 [5].

$$d_i = \sqrt{\sum_{i=1}^p (X_{2i} - X_{1i})^2} \quad (4)$$

Keterangan :

- = jarak
- X₂ = nilai data latih
- X₁ = nilai data uji
- p = dimensi data
- i = variable data

Hasil klasifikasi bumbu diperoleh dari proses pencocokan nilai antara data *training* dan data *testing*. Metode pengujian pada penelitian ini menggunakan akurasi dari evaluasi *confusion matrix*. Nilai akurasi diperoleh berdasarkan tingkat kedekatan nilai prediksi dengan nilai sebenarnya. Persamaan akurasi ditunjukkan pada Persamaan 5 [10].

$$Accuracy = \frac{tp+tn}{tp+fn+fp+tn} \quad (5)$$

Keterangan :

- tp (*true positive*) = jumlah data kelas sebenarnya positif dan kelas prediksi juga positif.
- fn (*false negative*) = jumlah data kelas sebenarnya positif dan kelas prediksi negatif
- fp (*false positive*) = jumlah data kelas sebenarnya negatif dan kelas prediksi positif.
- tn (*true negative*) = jumlah data kelas sebenarnya negatif dan kelas prediksi juga negatif

3. HASIL DAN PEMBAHASAN

Observasi dilakukan pada 2 hal yaitu variasi K pada KNN dan cara membagi data *testing* dan data *training*, Hasil percobaan pertama menggunakan pembagian data *training* dan data *testing* sebesar 66,7%: 33,33% dengan nilai K=1, K=3 dan K=5. Nilai K yang dipilih adalah ganjil karena jumlah kelas dalam klasifikasinya adalah genap[10]. Hasil percobaan menggunakan Bahasa pemrograman python dengan variasi nilai K dan cara pembagian data ditunjukkan pada Gambar 6 sampai dengan Gambar 14.

```
[INFO] features matrix: 2.3MB
[HASIL] Evaluasi dengan KNN
```

	precision	recall	f1-score	support
jahe	0.66	0.73	0.70	67
kencur	0.67	0.68	0.67	71
kunyit	0.79	0.79	0.79	57
lengkuas	0.79	0.70	0.74	69
accuracy			0.72	264
macro avg	0.73	0.72	0.72	264
weighted avg	0.72	0.72	0.72	264

Gambar 6. KNN dengan K=1 dengan Pembagian Data 66,7%: 33,33%

```
[INFO] features matrix: 2.3MB
[HASIL] Evaluasi dengan KNN
```

	precision	recall	f1-score	support
jahe	0.49	0.73	0.59	67
kencur	0.66	0.54	0.59	71
kunyit	0.80	0.70	0.75	57
lengkuas	0.70	0.57	0.62	69
accuracy			0.63	264
macro avg	0.66	0.63	0.64	264
weighted avg	0.66	0.63	0.63	264

Gambar 7. KNN dengan K=3 dengan Pembagian 66,7%: 33,33%

```
[INFO] features matrix: 2.3MB
[HASIL] Evaluasi dengan KNN
```

	precision	recall	f1-score	support
jahe	0.53	0.76	0.63	67
kencur	0.65	0.61	0.63	71
kunyit	0.82	0.70	0.75	57
lengkuas	0.74	0.57	0.64	69
accuracy			0.66	264
macro avg	0.68	0.66	0.66	264
weighted avg	0.68	0.66	0.66	264

Gambar 8. KNN dengan K=5 dengan Pembagian Data 66,7%: 33,33%

Hasil percobaan kedua menggunakan pembagian data *training* dan data *testing* sebesar 75%: 25% dengan nilai K=1 K=3 dan K=5 ditunjukkan pada gambar 9, Gambar 10, dan Gambar 11.

```
[INFO] features matrix: 2.3MB
[HASIL] Evaluasi dengan KNN
```

	precision	recall	f1-score	support
jahe	0.60	0.80	0.69	45
kencur	0.66	0.66	0.66	53
kunyit	0.81	0.76	0.78	45
lengkuas	0.82	0.65	0.73	57
accuracy			0.71	200
macro avg	0.72	0.72	0.71	200
weighted avg	0.73	0.71	0.71	200

Gambar 9. KNN dengan K=1 dengan Pembagian Data 66,7%: 33,33%

```
[INFO] features matrix: 2.3MB
[HASIL] Evaluasi dengan KNN
```

	precision	recall	f1-score	support
jahe	0.46	0.80	0.58	45
kencur	0.63	0.51	0.56	53
kunyit	0.81	0.67	0.73	45
lengkuas	0.78	0.56	0.65	57
accuracy			0.62	200
macro avg	0.67	0.63	0.63	200
weighted avg	0.67	0.62	0.63	200

Gambar 10. KNN dengan K=3 dengan Pembagian Data 75%:25%

```
[INFO] features matrix: 2.3MB
[HASIL] Evaluasi dengan KNN
```

	precision	recall	f1-score	support
jahe	0.47	0.76	0.58	45
kencur	0.60	0.51	0.55	53
kunyit	0.81	0.64	0.72	45
lengkuas	0.70	0.56	0.62	57
accuracy			0.61	200
macro avg	0.64	0.62	0.62	200
weighted avg	0.64	0.61	0.61	200

Gambar 11. KNN dengan K=5 dengan Pembagian Data 75%:25%

Hasil percobaan ketiga menggunakan pembagian data *training* dan data *testing* sebesar 90%:10% dengan nilai K=3 ditunjukkan pada gambar 12, gambar 3, dan gambar 14.

```
[INFO] features matrix: 2.3MB
[HASIL] Evaluasi dengan KNN
```

	precision	recall	f1-score	support
jahe	0.65	0.89	0.76	19
kencur	0.76	0.73	0.74	22
kunyit	0.79	0.79	0.79	19
lengkuas	0.93	0.65	0.76	20
accuracy			0.76	80
macro avg	0.78	0.77	0.76	80
weighted avg	0.78	0.76	0.76	80

Gambar 12. KNN dengan K=1 dengan Pembagian Data 90%:10%

```
[INFO] features matrix: 2.3MB
[HASIL] Evaluasi dengan KNN
```

	precision	recall	f1-score	support
jahe	0.52	0.89	0.65	19
kencur	0.67	0.55	0.60	22
kunyit	0.94	0.79	0.86	19
lengkuas	0.85	0.55	0.67	20
accuracy			0.69	80
macro avg	0.74	0.69	0.69	80
weighted avg	0.74	0.69	0.69	80

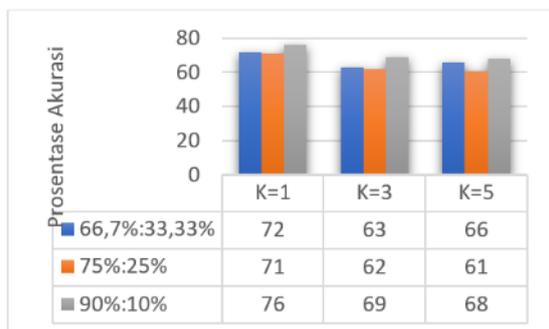
Gambar 13. KNN dengan K=3 dengan Pembagian Data 90%:10%

```
[INFO] features matrix: 2.3MB
[HASIL] Evaluasi dengan KNN
```

	precision	recall	f1-score	support
jahe	0.45	0.68	0.54	19
kencur	0.59	0.45	0.51	22
kunyit	0.75	0.63	0.69	19
lengkuas	0.72	0.65	0.68	20
accuracy			0.60	80
macro avg	0.63	0.61	0.61	80
weighted avg	0.63	0.60	0.60	80

Gambar 14. KNN dengan K=3 dengan Pembagian Data 90%:10%

Adapun hasil akurasi rata rata sesuai dengan nilai K ditunjukkan pada gambar 15.



Gambar 15. KNN dengan K=5 dengan Pembagian Data 90%:10%

Dari hasil perbandingan maka disimpulkan bahwa pemilihan K=3 atau K=5 memiliki rata rata akurasi yang hampir sama pada semua cara split data training dan data testing yaitu 64,66%: 65%. Pada K=1 memiliki akurasi yang cukup tinggi dibandingkan K sebelumnya yaitu 73% Disisi lain, dari cara pemilihan pembagian antara data *training* dan data testing, pembagian persentase data *training* dan data testing dengan persentase 90%:10% memiliki akurasi yang paling tinggi diikuti 66,7%: 33,33% kemudian 75%:25%. Klasifikasi bumbu dapur menggunakan KNN memiliki akurasi yang cukup baik sekitar 73% dengan nilai K=1. Akurasi ini harus dinaikkan dengan mencoba metode lain

yang lebih akurat. Akurasi dalam klasifikasi sangat berperan penting dalam implementasi. Selain itu, pemakaian faktor warna saja atau RGB dalam klasifikasi memiliki kerugian yang cukup besar. Fitur nya harus diperbanyak sehingga akurasinya bisa meningkat.

4. KESIMPULAN DAN SARAN

Kesimpulan dalam penelitian ini adalah metode klasifikasi bumbu dapur menggunakan KNN dengan fitur RGB memiliki akurasi yang sedang yaitu sekitar 73%. Pembagian data training dan data testing 90%:10% memiliki akurasi tertinggi pada semua nilai K.

5. REFERENSI

- [1] W. Setiawan and F. Damayanti, "Klasifikasi Citra Retina Menggunakan K-Nearest," in *Seminar Nasional Sains dan Teknologi*, 2016, no. November, pp. 1–6.
- [2] S. M. Ayyad, A. I. Saleh, and L. M. Labib, "Gene expression cancer classification using modified K-Nearest Neighbors technique," *BioSystems*, vol. 176, no. December 2018, pp. 41–51, 2019, doi: 10.1016/j.biosystems.2018.12.009.
- [3] A. R. Widiarti, "K-nearest neighbor performance for Nusantara scripts image transliteration," *J. Teknol. dan Sist. Komput.*, vol. 8, no. 2, pp. 150–156, 2020, doi: 10.14710/jtsiskom.8.2.2020.150-156.
- [4] S. H. Wardani, T. Rismawan, and S. Bahri, "Aplikasi Klasifikasi Jenis Tumbuhan Mangrove Berdasarkan Karakteristik Morfologi Menggunakan Metode K-Nearest Neighbor (KNN) Berbasis Web," *Coding J. Komput. dan Apl.* Untan, vol. 04 (3), no. 3, pp. 9–21, 2016.
- [5] E. Budianita, J. Jasril, and L. Handayani, "Implementasi Pengolahan Citra dan Klasifikasi K-Nearest Neighbour Untuk Membangun Aplikasi Pembeda Daging Sapi dan Babi Berbasis Web," *J. Sains dan Teknol. Ind.*, vol. 12, no. Vol 12, No 2 (2015): Juni 2015, pp. 242–247, 2015, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/sitekin/article/view/1005>.
- [6] B. S. Hutagaol, Y. A. Sari, and P. P. Adikara, "Ekstraksi Fitur RGB Color Channel dan Simple Morphological Shape Descriptors dari Citra Makanan untuk Pencarian Resep Makanan," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2923–2928, 2019.
- [7] R. Arian, A. Hariri, A. Mehridehnavi, A. Fassihi, and F. Ghasemi, "Protein kinase inhibitors' classification using K-Nearest neighbor algorithm," *Comput. Biol. Chem.*,

- vol. 86, p. 107269, 2020, doi: 10.1016/j.combiolchem.2020.107269.
- [8] A. Rosebrock, *Deep Learning for Computer Vision With Python*, 3rd ed. United States of America: PyImageSearch, 2018.
- [9] O. S. Y. Prakasa and K. M. Lhaksana, "Klasifikasi Teks Dengan Menggunakan Algoritma K-nearest Neighbor Pada Kasus Kinerja Pemerintah Di Twitter," *eProceedings Eng.*, vol. 5, no. 3, pp. 8237–8248, 2018.
- [10] L. Farokhah, "Implementasi K-Nearest Neighbor Untuk Klasifikasi Bunga Implementation of K-Nearest Neighbor for Flower Classification With Extraction of Rgb Color Features," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 1129–1136, 2020, doi: 10.25126/jtiik.202072608.