

Pengenalan Tulisan Tangan Huruf Latin Bersambung Menggunakan Local Binary Pattern dan K-Nearest Neighbor

Vivin Oktavia ⁽¹⁾, Novan Wijaya ^{(2)*}

¹ Teknik Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang, Palembang

² Magister Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang, Palembang

e-mail : vivinoktavia@mhs.mdp.ac.id, novan.wijaya@mdp.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 23 Juli 2022, direvisi 7 September 2022, diterima 11 September 2022, dan dipublikasikan 25 September 2022.

Abstract

There are 26 Latin letters in Indonesia, 5 of which are vowels and 21 consonants. This study will translate handwriting with a Latin object using the K-Nearest Neighbor method with the Local Binary Pattern extension. The research is being done with a focus on experimentation using a few methods that have already been discussed. Concatenated Latin letters have a few variations that depend on the work's author, so research will be conducted to identify cursive Latin letters based on these variations. Each of the 30 respondents wrote 26 capital letters and 26 lowercase letters on paper, which was then scanned to provide the image data. Black, blue, and red pens were used to write by every ten responders. The recognition procedure is broken into two halves, capital and non-capital letter recognition using 780 picture datasets each. In the study, k-fold cross-validation is used, with $k = 6$. The best value was reached at $k = 7$ with 29.49 percent accuracy, 33.88 percent precision, recall 33.46 percent, and F1-score 27.65 percent according to the research utilizing KNN with values $k = 3, 5, \text{ and } 7$. and for recognizing non-capital characters, the best result was found at $k=3$ with accuracy, precision, recall, and F1-score of 26.28, 27.27, and 22.7%, respectively.

Keywords: Latin Letters, Cursive, K-Nearest Neighbor, Local Binary Pattern, K-Fold Cross Validation, Handwritten

Abstrak

Huruf Latin yang dimiliki Indonesia berjumlah 26 huruf, di antaranya 5 huruf vokal dan 21 huruf konsonan. Dalam penelitian ini dilakukan pengenalan tulisan tangan dengan objek huruf Latin bersambung menggunakan metode K-Nearest Neighbor dengan ekstraksi ciri Local Binary Pattern. Penelitian yang dilakukan bersifat eksperimen dengan menggunakan beberapa metode yang telah disebutkan sebelumnya. Huruf Latin bersambung memiliki beberapa variasi tergantung dari penulis itu sendiri sehingga berdasarkan variasi tersebut dilakukan penelitian dalam mengidentifikasi huruf Latin bersambung. *Dataset* citra merupakan hasil *scan* tulisan tangan di atas kertas yang ditulis oleh 30 responden yang masing-masing responden menulis 26 huruf kapital dan 26 huruf non-kapital. Tiap 10 responden menulis menggunakan pena berwarna hitam, biru, dan merah. Proses pengenalan dibagi menjadi 2 yaitu pengenalan huruf kapital dengan 780 *dataset* citra dan non kapital dengan 780 *dataset* citra. Penelitian menggunakan *k-fold cross validation* dengan nilai k yaitu 6. Hasil dari penelitian menggunakan KNN dengan nilai 3, 5, dan 7, pada pengenalan huruf kapital didapatkan nilai terbaik pada $k=7$ dengan *accuracy* 29.49%, *precision* 33.88%, *recall* 33.46%, dan *F1-score* 27.65% dan pada pengenalan huruf non-kapital didapatkan nilai terbaik pada $k=3$ dengan *accuracy* 26.28%, *precision* 27.27%, *recall* 26.27%, dan *F1-score* 22.7%.

Kata Kunci: Huruf Latin, Huruf Latin Bersambung, K-Nearest Neighbor, Local Binary Pattern, K-Fold Cross Validation, Tulisan Tangan



1. PENDAHULUAN

Huruf merupakan bentuk, goresan, atau simbol dari suatu sistem tulisan. Huruf Latin merupakan salah satu aksara yang paling banyak digunakan dan merupakan aksara yang pertama kali dipakai oleh orang Romawi. Huruf Latin bersambung adalah suatu bentuk tulisan tangan yang tiap hurufnya ditulis secara terhubung. Huruf Latin sering ditulis oleh anak-anak sekolah dasar. Hal ini bermanfaat untuk perkembangan kemampuan motorik anak dan perkembangan otak (Maharani et al., 2019).

Kemajuan teknologi yang semakin hari mengalami perkembangan yang pesat, dapat membawa kemajuan yang sangat berarti dalam berbagai aspek kehidupan masyarakat. Penelitian mengenai pengenalan huruf (Masrani et al., 2018), angka (Prihatiningsih et al., 2019), dan tulisan tangan (Anggraeny et al., 2020) yang beragam sudah sering dilakukan. Salah satu metode untuk ekstraksi fitur tekstur adalah Local Binary Pattern. LBP memiliki kelebihan dapat mendeskripsikan karakter tekstur pada permukaan (Rahayu et al., 2021). Lalu metode yang dapat digunakan untuk melakukan pengenalan objek adalah K-Nearest Neighbor. Beberapa penelitian yang telah dilakukan mengenai tulisan tangan ini, di antaranya tulisan tangan dengan huruf Latin, tulisan tangan aksara Jawa dan sebagainya. Penelitian yang telah dilakukan sebelumnya oleh (Purbayanti, 2018) menggunakan *dataset* berupa huruf Latin, K-Nearest Neighbor, dan deteksi sobel menghasilkan tingkat akurasi tertinggi dengan jumlah data yang bernilai benar adalah 84 dari 90 data sebesar 93,334% dan akurasi terendah dengan jumlah data yang bernilai benar adalah 39 dari 72 sebesar 54,167%.

Penelitian yang dilakukan (Ilham & Rochmawati, 2020) menggunakan metode K-Nearest Neighbor menggunakan deteksi tepi Sobel untuk memperoleh bagian detail dengan perhitungan jarak dengan 3 skenario. Skenario pertama menggunakan data *training* sebanyak 40 data dan data *testing* sebanyak 20 data dengan menghasilkan akurasi sebesar 40%, skenario kedua menggunakan data *training* sebanyak 60 data dan data *testing* sebanyak 20 data dengan mencapai akurasi sebesar 55%, skenario ketiga menggunakan 100 data *training* dan 20 data *testing* lalu diperoleh akurasi sebesar 85%. Penelitian tentang unjuk kerja K-Nearest Neighbors terhadap Penggalan Karakter Jawa yang menggunakan Local Binary Pattern yang berguna untuk mengekstrak ciri unik dari citra tulisan aksara Jawa. Sementara KNN digunakan untuk menentukan kelas dari citra aksara Jawa berdasarkan hasil ekstraksi ciri LBP. Pengujian dilakukan dengan menggunakan 160 citra yang dibagi menjadi 40 citra uji dan 120 citra latih. Hasil pengujian dievaluasi menggunakan *confusion matrix* untuk menentukan akurasi dari kombinasi KNN dan LBP dengan akurasi tertinggi mencapai 82.5% di mana parameter yang digunakan adalah *cell size* berukuran 64x64 dan nilai $k=3$ (Bimantoro et al., 2021).

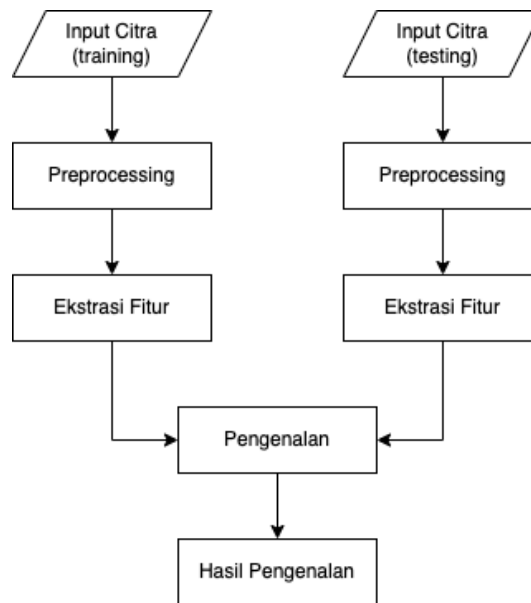
Berdasarkan studi literatur yang ada, sudah banyak penelitian mengenai pengenalan tulisan tangan dengan berbagai metode dan berbagai objek seperti huruf, angka, aksara Korea, aksara Jepang, aksara Jawa, dan lainnya. Banyaknya variasi huruf Latin bersambung yang berbeda dari penulisnya diterjemahkan ke dalam komputer dan dilakukan proses pengenalan dari huruf Latin bersambung tersebut. Belum banyak penelitian pengenalan tulisan tangan dengan objek huruf Latin terutama huruf Latin yang bersifat bersambung. Beberapa penelitian yang telah dilakukan terkait tulisan tangan dapat disimpulkan belum didapatkannya penelitian yang bersifat secara spesifik membahas mengenai tulisan tangan bersambung baik kapital dan non kapital. Penelitian yang dilakukan bersifat eksperimen menggunakan *local binary pattern* sebagai ekstraksi ciri yang di mana proses untuk mendapatkan ciri unik dari pola citra yang dijadikan data input pada pengenalan citra dengan menggunakan K-Nearest Neighbor sebagai pengenalan huruf Latin bersambung kapital dan non kapital.

2. METODE PENELITIAN

Tulisan tangan adalah salah satu hal unik yang dapat dihasilkan oleh manusia selain tanda tangan (Andana et al., 2018). Beberapa penelitian menunjukkan bahwa perangkat lunak dapat mengenali tulisan tangan setiap orang. *Handwriting Recognition* atau pengenalan tulisan tangan adalah suatu proses di mana komputer menerjemahkan tulisan tangan kedalam teks komputer



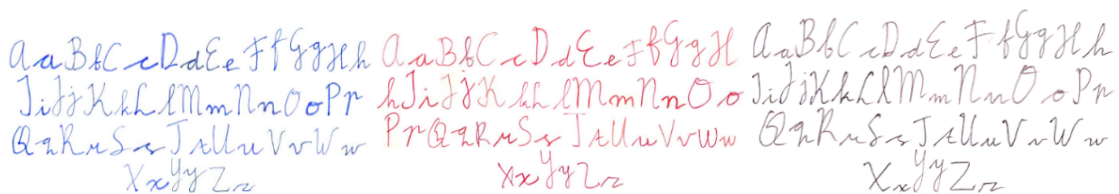
(Cahyani et al., 2018). Huruf Latin adalah huruf yang biasa digunakan dalam tulisan tangan yang ditulis berangkai-rangkai. Huruf Latin bersambung merupakan suatu bentuk atau gores dari suatu sistem tulisan di mana tulisan ditulis secara terhubung (Aranta et al., 2020). Pengenalan tulisan tangan bersambung pada penelitian ini memiliki beberapa proses seperti yang ditunjukkan pada Gambar 1.



Gambar 1 Gambaran Umum

2.1 Dataset

Dataset diambil dari 30 orang responden yang di mana tiap 10 responden akan menulis menggunakan pena berwarna hitam, biru, dan merah. Setiap responden akan menulis 52 huruf (26 huruf kapital dan 26 huruf non-kapital) sehingga total keseluruhan data adalah 1560 foto (Gambar 2). Proses pengumpulan dataset tulisan tangan berupa foto *scan* dan dibagi menjadi 2 kelompok yaitu 780 huruf Latin bersambung kapital dan 780 huruf Latin bersambung non-kapital. *Dataset* tersebut akan menjadi *dataset* latih dan *dataset* uji.

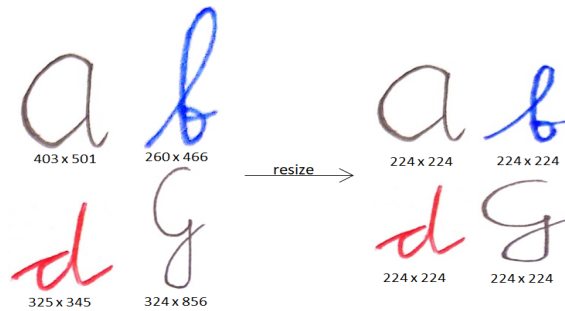


Gambar 2 *Dataset* Tiap Jenis Huruf Latin Bersambung

2.2 Preprocessing

Pada tahap *preprocessing* (pengolah awal) bertujuan untuk mendapatkan informasi dari citra serta meningkatkan kualitas citra yang di mana pada tahap ini meliputi *resize* dan *grayscale* dari LBP. Citra awal diberi label secara manual sesuai klasifikasi yang berukuran bervariasi, lalu semua data akan melewati proses *resize* sehingga semua ukuran citra berukuran sama yaitu 224×224 piksel. Proses *resize* dapat diilustrasikan pada Gambar 3.

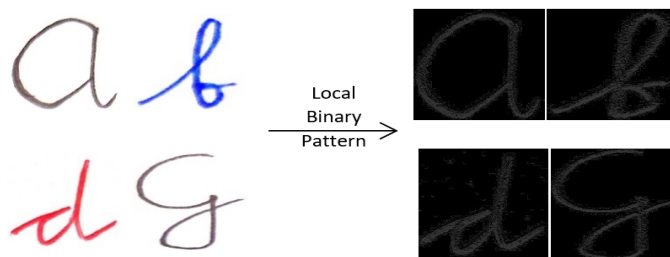




Gambar 3 Dataset Resize

2.3 Ekstraksi Fitur

Untuk ekstraksi fitur pada penelitian ini menggunakan ekstraksi ciri tekstur *Local Binary Pattern*. Local Binary Pattern merupakan metode yang dapat menghasilkan tekstur dari sebuah citra *grayscale* yang bekerja dengan memanfaatkan nilai piksel ketetanggaan yang tersebar melingkar (*circular neighborhood*) dengan berbagai jenis ukuran yang direpresentasikan dalam bentuk *matrix* (Al Rivan et al., 2020). Setelah *dataset* *resize*, kemudian dilakukan ekstraksi ciri *Local Binary Pattern*. Proses LBP dapat diilustrasikan pada Gambar 4.



Gambar 4 Local Binary Pattern

Dengan menerapkan operator dasar LBP yang berukuran 3x3, diambil nilai tengah sebagai ambang batas (*threshold*) dan dibandingkan dengan nilai sekelilingnya. Jika piksel sekeliling lebih besar dibandingkan dengan ambang batas maka akan bernilai 1, dan jika nilai piksel sekeliling lebih kecil dibandingkan dengan ambang batas maka akan bernilai 0. Setelah didapat nilai biner, maka akan dikali dengan nilai bobot 2^p yang adalah urutan piksel tetangga dimulai dari pojok kiri atas hingga pojok kanan bawah dengan mengelilingi piksel tengah yang dimulai dari 0 sampai 7. Hasil dari citra *grayscale* yang dilakukan ekstraksi ciri yang dapat dilihat pada Gambar 5 dan Gambar 6.

	0	1	2	3	4	5	6	7	8	9	...	49	50	51	52	53	54	55	56	57	58	
0	24769.0	618.0	527.0	1654.0	799.0	48.0	1388.0	881.0	45.0	36.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11008.0
1	34526.0	353.0	340.0	1007.0	333.0	28.0	406.0	298.0	38.0	32.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8728.0
2	14705.0	1185.0	770.0	2795.0	1455.0	93.0	1660.0	1438.0	117.0	94.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	17107.0
3	27493.0	630.0	501.0	1668.0	746.0	50.0	1357.0	677.0	57.0	37.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10509.0
4	10838.0	1327.0	790.0	2775.0	1558.0	134.0	1408.0	1598.0	169.0	121.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20868.0
...
775	39530.0	57.0	51.0	169.0	92.0	30.0	723.0	81.0	31.0	36.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6130.0
776	41644.0	72.0	56.0	188.0	64.0	31.0	600.0	73.0	31.0	29.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4110.0
777	38628.0	80.0	61.0	259.0	128.0	44.0	445.0	129.0	33.0	46.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7216.0
778	29336.0	240.0	238.0	680.0	423.0	76.0	845.0	604.0	92.0	83.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11197.0
779	36091.0	160.0	110.0	343.0	251.0	36.0	587.0	215.0	51.0	63.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8234.0

780 rows x 59 columns

Gambar 5 Hasil Ekstraksi Ciri LBP Huruf Kapital Latin Bersambung



	0	1	2	3	4	5	6	7	8	9	...	49	50	51	52	53	54	55	56	57	58	
0	34316.0	248.0	165.0	608.0	249.0	18.0	515.0	227.0	21.0	27.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10034.0
1	29005.0	497.0	472.0	1278.0	569.0	43.0	1449.0	578.0	43.0	24.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9487.0
2	32432.0	288.0	191.0	396.0	199.0	10.0	466.0	303.0	30.0	30.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12258.0
3	30993.0	400.0	192.0	398.0	256.0	10.0	454.0	223.0	16.0	41.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13070.0
4	23782.0	851.0	634.0	2002.0	854.0	70.0	1209.0	833.0	91.0	70.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13193.0
...
775	37892.0	144.0	75.0	204.0	118.0	22.0	830.0	156.0	17.0	27.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6822.0
776	37652.0	134.0	132.0	307.0	193.0	46.0	670.0	167.0	33.0	47.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7084.0
777	39225.0	64.0	76.0	100.0	114.0	34.0	602.0	96.0	18.0	34.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6725.0
778	38962.0	111.0	94.0	180.0	123.0	29.0	628.0	122.0	31.0	33.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6293.0
779	37655.0	135.0	94.0	312.0	202.0	28.0	677.0	155.0	51.0	36.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6908.0

780 rows x 59 columns

Gambar 6 Hasil Ekstraksi Ciri LBP Huruf Non-Kapital Latin Bersambung

Dataset citra yang sudah dilakukan ekstraksi fitur akan dinormalisasikan terlebih dahulu sebelum dilanjutkan ke tahap pengenalan huruf. Gambar 7 dan Gambar 8 merupakan hasil normalisasi terhadap ekstraksi fitur LBP huruf kapital dan huruf non-kapital Latin bersambung.

	0	1	2	3	4	5	6	7	8	9	...	50	51	52	53	54	55	56	57	58	target		
0	0.901943	0.022504	0.019190	0.060229	0.029095	0.001748	0.050543	0.032081	0.001639	0.001311	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.400847	9
1	0.968369	0.009901	0.009536	0.028244	0.009340	0.000785	0.011387	0.008358	0.001066	0.000898	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.244799	23
2	0.635821	0.051238	0.033294	0.120851	0.062912	0.004021	0.071776	0.062177	0.005059	0.004064	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.739679	19
3	0.928217	0.021270	0.016915	0.056315	0.025186	0.001688	0.045815	0.022857	0.001924	0.001249	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.354804	25
4	0.450885	0.055206	0.032866	0.115446	0.064816	0.005575	0.058576	0.066480	0.007031	0.005034	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.868156	24
...
775	0.987657	0.001424	0.001274	0.004222	0.002299	0.000750	0.018064	0.002024	0.000775	0.000899	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.153158	6
776	0.994748	0.001720	0.001338	0.004491	0.001529	0.000740	0.014332	0.001744	0.000740	0.000693	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.098175	5
777	0.982663	0.002035	0.001552	0.006589	0.003256	0.001119	0.011320	0.003282	0.000839	0.001170	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.183569	17
778	0.931225	0.007618	0.007555	0.021586	0.013427	0.002413	0.026823	0.019173	0.002920	0.002635	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.355431	1
779	0.974253	0.004319	0.002969	0.009259	0.006776	0.000972	0.015846	0.005804	0.001377	0.001701	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.222271	17

780 rows x 60 columns

Gambar 7 Hasil Ekstraksi LBP Huruf Kapital yang Telah Dinormalisasi

	0	1	2	3	4	5	6	7	8	9	...	50	51	52	53	54	55	56	57	58	target		
0	0.959018	0.006931	0.004611	0.016992	0.006959	0.000503	0.014393	0.006344	0.000587	0.000755	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.280417	19
1	0.944450	0.016183	0.015369	0.041614	0.018528	0.001400	0.047182	0.018821	0.001400	0.000781	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.308912	9
2	0.934855	0.008302	0.005506	0.011415	0.005736	0.000288	0.013432	0.008734	0.000865	0.000865	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.353338	23
3	0.920691	0.011883	0.005704	0.011823	0.007605	0.000297	0.013487	0.006625	0.000475	0.001218	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.388263	22
4	0.867197	0.031031	0.023118	0.073002	0.031141	0.002563	0.044086	0.030375	0.003318	0.002563	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.481075	24
...
775	0.983452	0.003737	0.001947	0.005295	0.003063	0.000571	0.021542	0.004049	0.000441	0.000701	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.177059	16
776	0.982126	0.003495	0.003443	0.008008	0.005034	0.001200	0.017476	0.004356	0.000861	0.001226	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.184781	6
777	0.985206	0.001607	0.001909	0.002512	0.002863	0.000854	0.015120	0.002411	0.000452	0.000854	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.168910	6
778	0.986669	0.002811	0.002380	0.004558	0.003115	0.000734	0.015903	0.003090	0.000785	0.000836	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.159363	6
779	0.982850	0.003524	0.002454	0.008144	0.005272	0.000731	0.017671	0.004046	0.001331	0.000940	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.180309	1

780 rows x 60 columns

Gambar 8 Hasil Ekstraksi LBP Huruf Non-Kapital yang Telah Dinormalisasi



2.4 Tahap Pengenalan

Algoritma K-Nearest Neighbor adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut (Hidayat & Minati, 2019). Algoritma klasifikasi K-Nearest Neighbor termasuk algoritma *supervised learning*, di mana hasil dari query instance yang baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN (Nikmatun & Waspada, 2019). Metode KNN (K-Nearest Neighbor) melakukan pengenalan dengan objek huruf Latin bersambung untuk mengklasifikasi 26 jenis huruf Latin dengan masing-masing huruf dibagi menjadi 2 tipe yaitu huruf kapital dan huruf non-kapital. Pengujian menggunakan K-Nearest Neighbor dengan nilai $k = 3, 5, \text{ dan } 7$.

Pada pendekatan metode *k-fold cross validation*, *dataset* dibagi menjadi sejumlah k buah partisi secara acak. Selanjutnya, dilakukan sejumlah k -kali eksperimen dengan masing-masing eksperimen menggunakan data partisi ke- k sebagai data *testing* dan menggunakan sisa partisi lainnya sebagai data *training* (Pangestu et al., 2020). Data *training* dan data *testing* dibagi menggunakan *k-fold cross validation* dengan $k = 6$. Hasil dari pengujian akan dihitung menggunakan *confusion matrix* yang di mana *matrix* berukuran 26×26 dikarenakan data kelas berjumlah 26 untuk mendapatkan nilai *precision*, *recall*, *F1-score* dan *accuracy*.

Confusion matrix digunakan untuk melihat performa dari suatu model yang telah dibuat yaitu *accuracy*, *precision*, *recall*, dan *F1-score* (Maulidah et al., 2020).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4)$$

Di mana *True Positive (TP)* merupakan jumlah data positif citra yang terklasifikasi benar oleh sistem, *True Negative (TN)* merupakan jumlah data negatif citra yang terklasifikasi benar oleh sistem, *False Positive (FP)* merupakan jumlah data positif citra yang terklasifikasi salah oleh sistem, dan *False Negative (FN)* merupakan jumlah data negatif citra yang terklasifikasi salah oleh sistem.

3. HASIL DAN PEMBAHASAN

3.1 Pembahasan Pengujian KNN Huruf Kapital

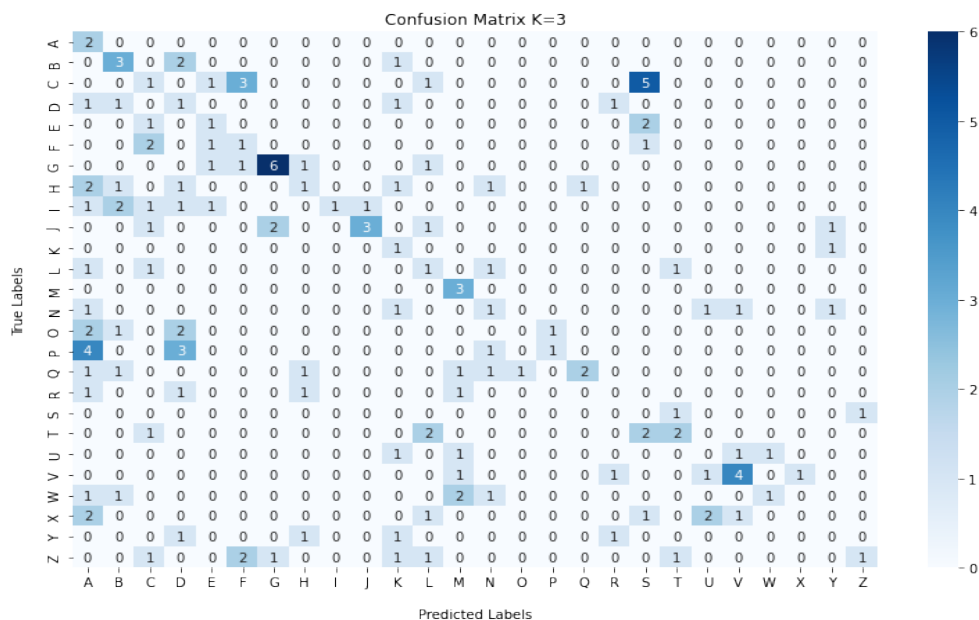
Proses pengujian huruf kapital Latin bersambung menggunakan K-Nearest Neighbor di mana pada penelitian ini nilai k yang digunakan adalah 3, 5, dan 7. Hasil pengujian pengenalan huruf kapital $k=3$, $k=5$, dan $k=7$ dapat dilihat pada Tabel 1 sampai 3.

Gambar 9 dirincikan hasil prediksi huruf G yang dapat dilihat melalui *confusion matrix*. Kelas G dapat dikenali sebanyak 6 data yang merupakan nilai *True Positive*. Adapula data bernilai G yang diprediksi bukan G yang merupakan *False Negative* sebanyak 4 dan data yang terprediksi G yang bukan data G yang merupakan *False Positive* sebanyak 3 dengan sisa nilai lain sebagai *True Negative*. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision* 67%, *recall* 60%, *F1-score* 63%, dan *accuracy* 95.51%.



Tabel 1 Hasil Pengujian Pengenalan Huruf Kapital K=3

Huruf	Precision	Recall	F1-Score	Accuracy
A	11%	100%	19%	89.1%
B	30%	50%	37%	93.59%
C	11%	9%	10%	88.46%
D	8%	20%	12%	90.38%
E	20%	25%	22%	95.51%
F	14%	20%	17%	93.59%
G	67%	60%	63%	95.51%
H	20%	12%	15%	92.95%
I	100%	12%	22%	95.51%
J	75%	38%	50%	96.15%
K	12%	50%	20%	94.87%
L	12%	20%	15%	92.95%
M	33%	100%	50%	96.15%
N	17%	17%	17%	93.59%
O	0%	0%	0%	95.51%
P	50%	11%	18%	94.23%
Q	67%	25%	36%	95.51%
R	0%	0%	0%	95.51%
S	0%	0%	0%	91.67%
T	40%	29%	33%	94.87%
U	0%	0%	0%	96.15%
V	57%	50%	53%	95.51%
W	50%	17%	25%	96.15%
X	0%	0%	0%	94.87%
Y	0%	0%	0%	95.51%
Z	50%	12%	20%	94.87%
Rata-Rata per Kelas	28.61%	26.04%	21.31%	94.18%



Gambar 9 Confusion Matrix Huruf Kapital K=3

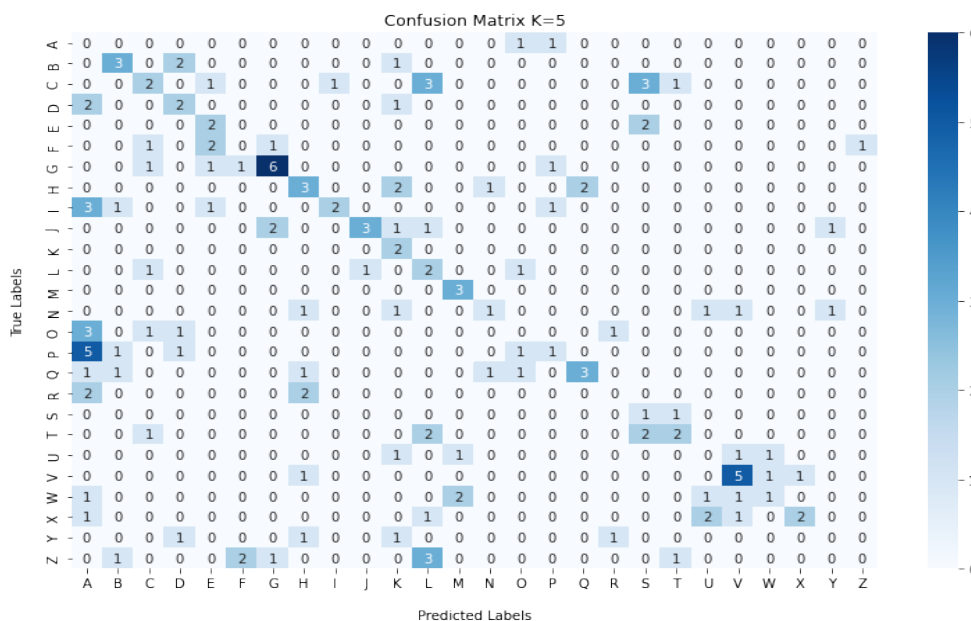
Gambar 10 dirincikan hasil prediksi huruf G yang dapat dilihat melalui *confusion matrix*. Kelas G dapat dikenali sebanyak 6 data yang merupakan nilai *True Positive*. Adapula data bernilai G yang diprediksi bukan G yang merupakan *False Negative* sebanyak 4 dan data yang terprediksi G yang



bukan data G yang merupakan *False Positive* sebanyak 4 dengan sisa nilai lain sebagai *True Negative*. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision*, *recall* dan *F1-score* masing-masing 60% dan *accuracy* 94.87%.

Tabel 2 Hasil Pengujian Pengenalan Huruf Kapital K=5

Huruf	Precision	Recall	F1-Score	Accuracy
A	0%	0%	0%	87.17%
B	43%	50%	46%	95.51%
C	29%	18%	22%	91.03%
D	29%	40%	33%	94.87%
E	29%	50%	36%	95.51%
F	0%	0%	0%	95.51%
G	60%	60%	60%	94.87%
H	33%	38%	35%	92.95%
I	67%	25%	36%	95.51%
J	75%	38%	50%	96.15%
K	20%	100%	33%	94.87%
L	17%	40%	24%	91.67%
M	50%	100%	67%	98.08%
N	33%	17%	22%	95.51%
O	0%	0%	0%	93.59%
P	25%	11%	15%	92.95%
Q	60%	38%	46%	95.51%
R	0%	0%	0%	96.15%
S	12%	50%	20%	94.87%
T	40%	29%	33%	94.87%
U	0%	0%	0%	94.87%
V	56%	62%	59%	95.51%
W	33%	17%	22%	95.51%
X	67%	29%	40%	96.15%
Y	0%	0%	0%	96.15%
Z	0%	0%	0%	94.23%
Rata-Rata per Kelas	29.92%	31.23%	26.88%	94.59%

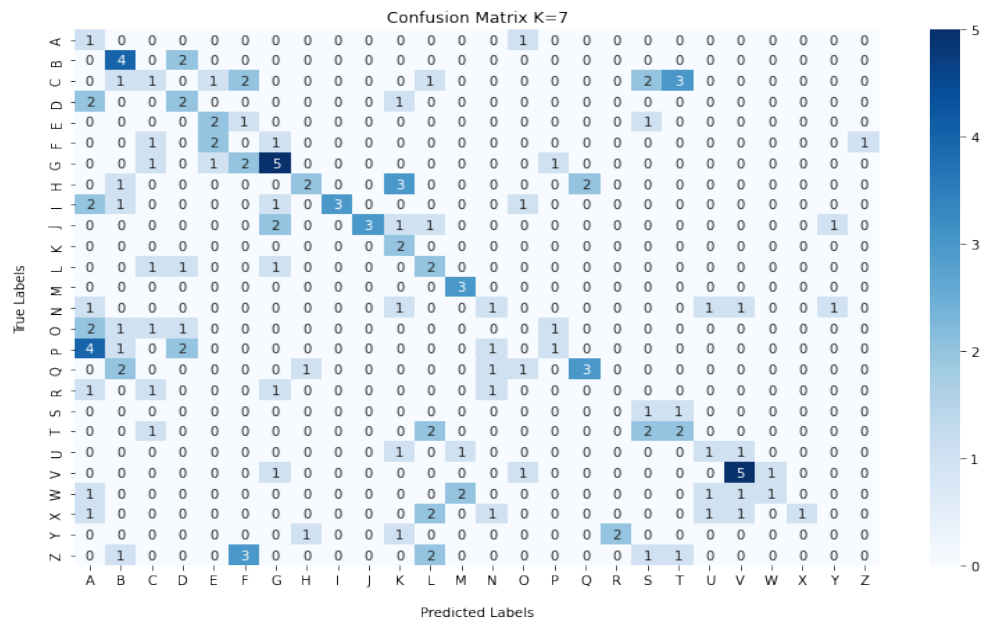


Gambar 10 Confusion Matrix Huruf Kapital K=5



Tabel 3 Hasil Pengujian Pengenalan Huruf Kapital K=7

Huruf	Precision	Recall	F1-Score	Accuracy
A	7%	50%	12%	90.38%
B	33%	67%	44%	93.59%
C	14%	9%	11%	89.74%
D	25%	40%	31%	94.23%
E	33%	50%	40%	96.15%
F	0%	0%	0%	91.67%
G	42%	50%	45%	92.31%
H	50%	25%	33%	94.87%
I	100%	38%	55%	96.79%
J	100%	38%	55%	96.79%
K	20%	100%	33%	94.87%
L	20%	40%	27%	92.95%
M	50%	100%	67%	98.08%
N	20%	17%	18%	94.23%
O	0%	0%	0%	93.59%
P	33%	11%	17%	93.59%
Q	60%	38%	46%	95.51%
R	0%	0%	0%	96.15%
S	14%	50%	22%	95.51%
T	29%	29%	29%	93.59%
U	25%	25%	25%	96.15%
V	56%	62%	59%	95.51%
W	50%	17%	25%	96.15%
X	100%	14%	25%	96.15%
Y	0%	0%	0%	96.15%
Z	0%	0%	0%	94.23%
Rata-Rata per Kelas	33.88%	33.46%	27.65%	94.57%



Gambar 11 Confusion Matrix Huruf Kapital K=7

Gambar 11 dirincikan hasil prediksi huruf G dan V yang dapat dilihat melalui *confusion matrix*. Kelas G dapat dikenali sebanyak 5 data yang merupakan nilai *True Positive*. Adapula data bernilai G yang diprediksi bukan G yang merupakan *False Negative* sebanyak 4 dan data yang terprediksi



G yang bukan data G yang merupakan *False Positive* sebanyak 7 dengan sisa nilai lain sebagai *True Negative*. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision* 42%, *recall* 50%, *F1-score* 45%, dan *accuracy* 92.31%.

Lalu kelas V dapat dikenali sebanyak 5 data yang merupakan nilai *True Positive*. Adapula data bernilai V yang diprediksi bukan V yang merupakan *False Negative* sebanyak 3 dan data yang terprediksi V yang bukan data V yang merupakan *False Positive* sebanyak 4 dengan sisa nilai lain sebagai *True Negative*. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision* 56%, *recall* 62%, *F1-score* 59%, dan *accuracy* 95.51%.

3.2 Pembahasan Pengujian KNN Non Huruf Kapital

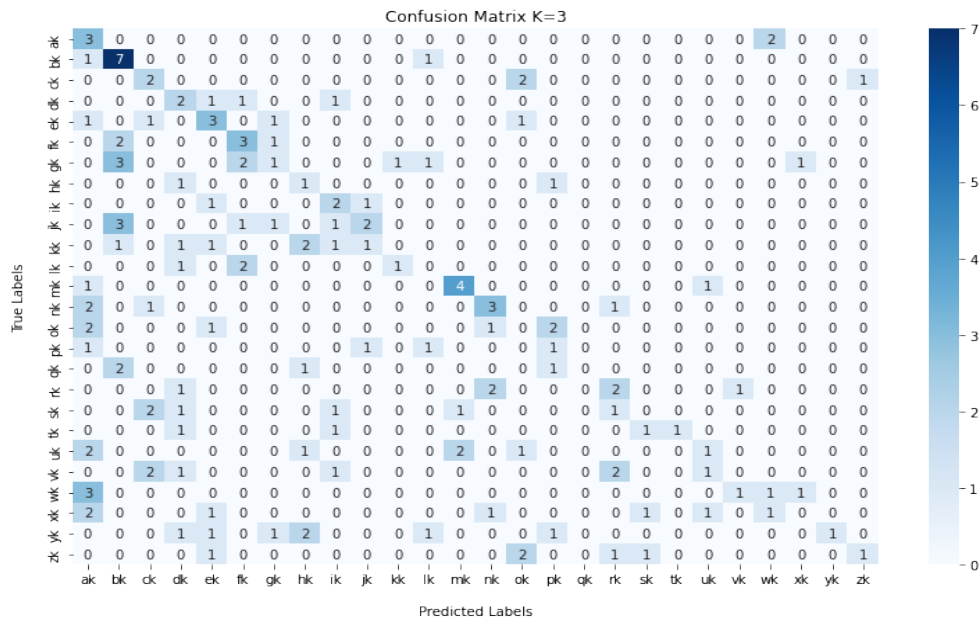
Proses pengujian pengenalan huruf non-kapital sama seperti proses pengenalan skenario pertama di mana nilai k yang digunakan adalah 3, 5, dan 7. Hasil pengenalan huruf non-kapital k=3, k=5, dan k=7 dapat dilihat pada Tabel 4 sampai 6.

Tabel 4 Hasil Pengujian Pengenalan Non Kapital K=3

Huruf	Precision	Recall	F1-Score	Accuracy
a	17%	60%	26%	89.1%
b	39%	78%	52%	91.67%
c	25%	40%	31%	94.23%
d	20%	40%	27%	92.95%
e	30%	43%	35%	92.95%
f	33%	50%	40%	94.23%
g	20%	11%	14%	92.31%
h	14%	33%	20%	94.87%
i	25%	50%	33%	94.87%
j	40%	25%	31%	94.23%
k	0%	0%	0%	94.23%
l	0%	0%	0%	94.87%
m	57%	67%	62%	96.79%
n	43%	43%	43%	94.87%
o	0%	0%	0%	92.31%
p	17%	25%	20%	94.87%
q	0%	0%	0%	97.43%
r	29%	33%	31%	94.23%
s	0%	0%	0%	94.23%
t	100%	25%	40%	98.08%
u	25%	14%	18%	94.23%
v	0%	0%	0%	94.23%
w	25%	17%	20%	94.87%
x	0%	0%	0%	94.23%
y	100%	12%	22%	95.51%
z	50%	17%	25%	96.15%
Rata-Rata per Kelas	27.27%	26.27%	22.7%	94.33%

Gambar 12 dirincikan hasil prediksi huruf b yang dapat dilihat melalui *confusion matrix*. Kelas b dapat dikenali sebanyak 7 data yang merupakan nilai *True Positive*. Adapula data bernilai b yang diprediksi bukan b yang merupakan *False Negative* sebanyak 2 dan data yang terprediksi b yang bukan data b yang merupakan *False Positive* sebanyak 11 dengan sisa nilai lain sebagai *True Negative*. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision* 39%, *recall* 78%, *F1-score* 52%, dan *accuracy* 91.67%.





Gambar 12 Confusion Matrix Huruf Kapital K=3

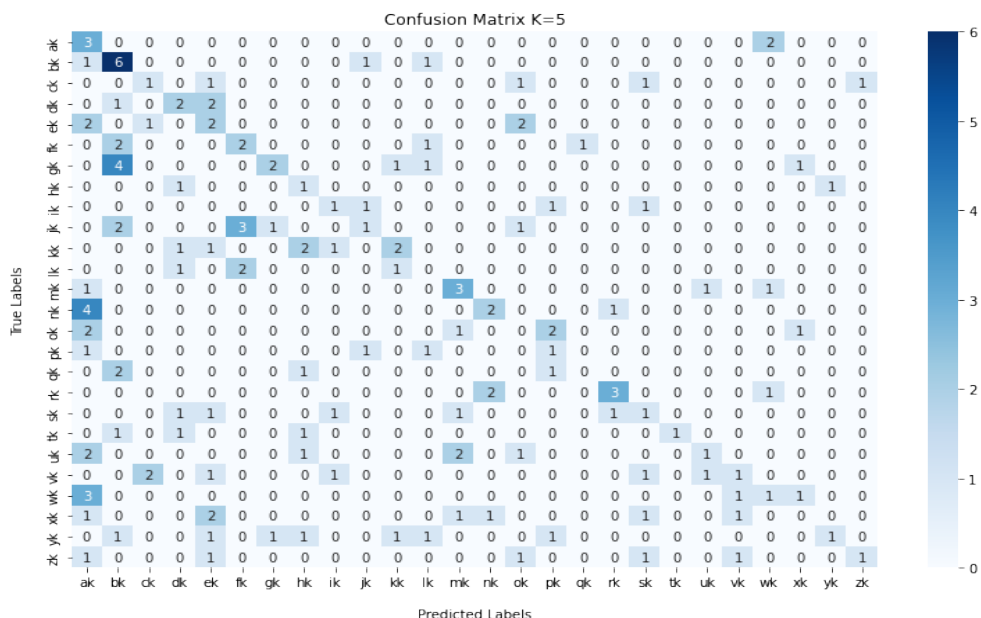
Tabel 5 Hasil Pengujian Pengenalan Non Kapital K=5

Huruf	Precision	Recall	F1-Score	Accuracy
a	14%	60%	23%	87.18%
b	32%	67%	43%	89.74%
c	25%	20%	22%	95.51%
d	29%	40%	33%	94.87%
e	17%	29%	21%	90.38%
f	29%	33%	31%	94.23%
g	50%	22%	31%	94.23%
h	14%	33%	20%	94.87%
i	25%	25%	25%	96.15%
j	25%	12%	17%	93.59%
k	40%	29%	33%	94.87%
l	0%	0%	0%	94.23%
m	38%	50%	43%	94.87%
n	40%	29%	33%	94.87%
o	0%	0%	0%	92.31%
p	17%	25%	20%	94.87%
q	0%	0%	0%	96.79%
r	60%	50%	55%	96.79%
s	17%	17%	17%	93.59%
t	100%	25%	40%	98.08%
u	33%	14%	20%	94.87%
v	25%	14%	18%	94.23%
w	20%	17%	18%	94.23%
x	0%	0%	0%	93.59%
y	50%	12%	20%	94.87%
z	50%	17%	25%	96.15%
Rata-Rata per Kelas	28.85%	24.61%	23.38%	94.23%

Gambar 13 dirincikan hasil prediksi huruf b yang dapat dilihat melalui *confusion matrix*. Kelas b dapat dikenali sebanyak 6 data yang merupakan nilai *True Positive*. Adapula data bernilai b yang diprediksi bukan b yang merupakan *False Negative* sebanyak 3 dan data yang terprediksi b yang



bukan data b yang merupakan *False Positive* sebanyak 13 dengan sisa nilai lain sebagai *True Negative*. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision* 32%, *recall* 67%, *F1-score* 43%, dan *accuracy* 89.74%.



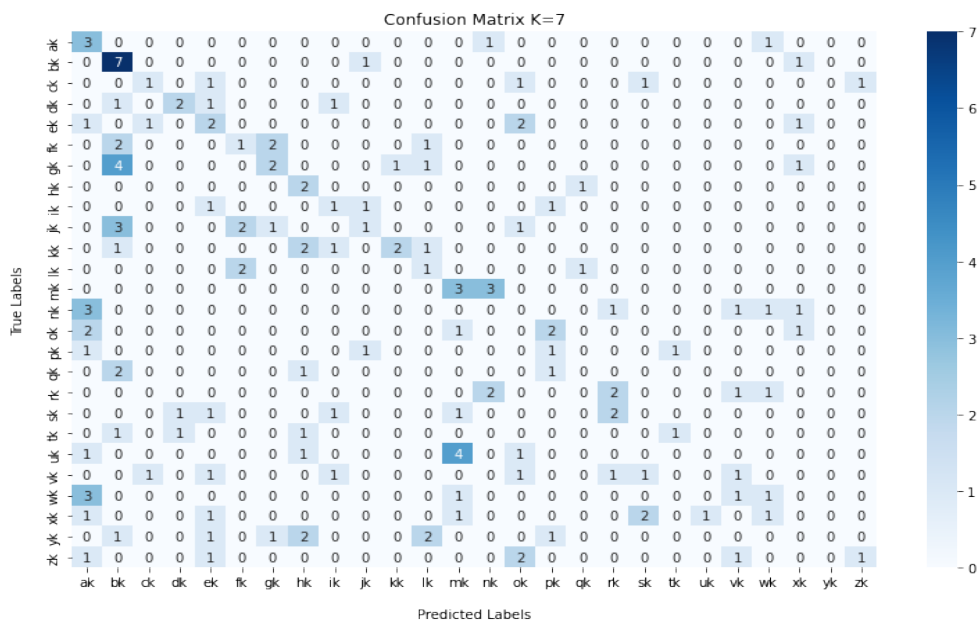
Gambar 13 *Confusion Matrix* Huruf Non Kapital K=5

Tabel 6 Hasil Pengujian Pengenalan Non Kapital K=7

Huruf	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Accuracy</i>
a	19%	60%	29%	90.38%
b	32%	78%	45%	89.1%
c	33%	20%	25%	96.15%
d	50%	40%	44%	96.79%
e	20%	29%	24%	91.67%
f	20%	17%	18%	94.23%
g	33%	22%	27%	92.95%
h	22%	67%	33%	94.87%
i	20%	25%	22%	95.51%
j	25%	12%	17%	93.59%
k	67%	29%	40%	96.15%
l	17%	25%	20%	94.87%
m	27%	50%	35%	92.95%
n	0%	0%	0%	91.67%
o	0%	0%	0%	91.02%
p	17%	25%	20%	94.87%
q	0%	0%	0%	96.15%
r	33%	33%	33%	94.87%
s	0%	0%	0%	93.59%
t	50%	25%	33%	97.43%
u	0%	0%	0%	94.87%
v	20%	14%	17%	93.59%
w	20%	17%	18%	94.23%
x	0%	0%	0%	92.31%
y	0%	0%	0%	94.87%
z	50%	17%	25%	96.15%
Rata-Rata per Kelas	22.12%	23.27%	20.19%	94.03%



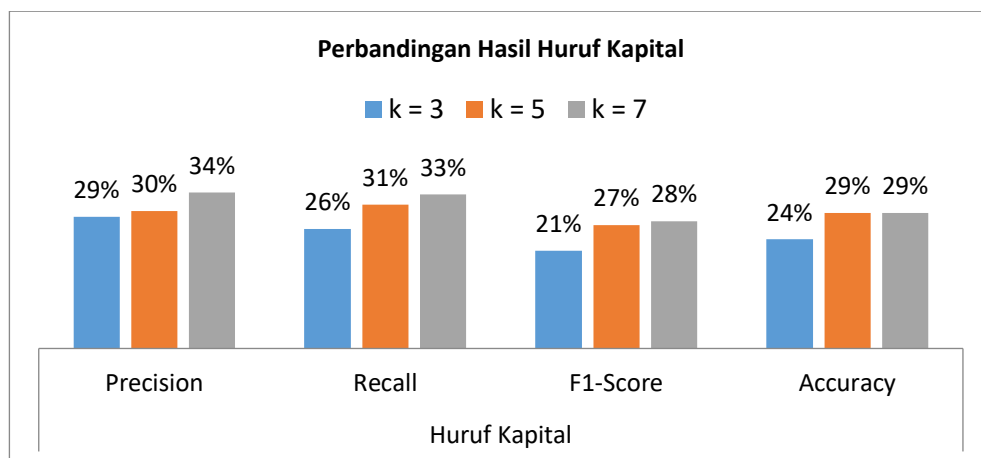
Gambar 14 dirincikan hasil prediksi huruf b yang dapat dilihat melalui *confusion matrix*. Kelas b dapat dikenali sebanyak 7 data yang merupakan nilai *True Positive*. Adapula data bernilai b yang diprediksi bukan b yang merupakan *False Negative* sebanyak 2 dan data yang terprediksi b yang bukan data b yang merupakan *False Positive* sebanyak 15 dengan sisa nilai lain sebagai *True Negative*. Dengan nilai TP, TN, FP, dan FN yang telah diketahui maka nilai *precision* 32%, *recall* 78%, *F1-score* 45%, dan *accuracy* 89.1%.



Gambar 14 Confusion Matrix Huruf Non Kapital K=7

3.3 Perbandingan Huruf Kapital dan Non Kapital

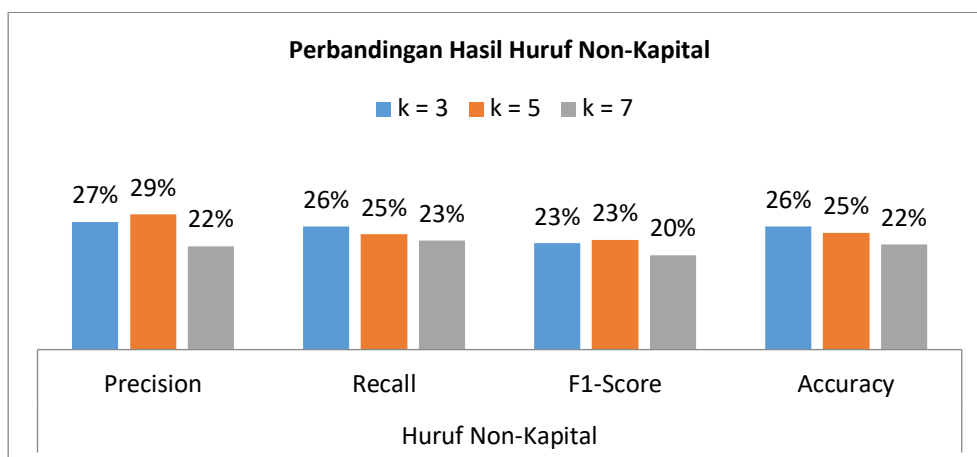
Gambar 15 pada bagian huruf kapital dapat disimpulkan bahwa nilai pengujian *accuracy* tertinggi pada huruf kapital yaitu pada k = 5 dan k = 7 dengan hasil 29,49%, sedangkan pada k = 3 hanya menghasilkan 23.72%. Selanjutnya nilai rata-rata *precision*, *recall*, dan *F1-score* tertinggi yaitu pada k = 7 dengan masing-masing 33.88%, 33.46%, dan 27.65%, sedangkan *Precision* pada k = 5 mendapatkan hasil 29.92% dan k = 3 menghasilkan 28.61%, *recall* pada k = 5 mendapatkan hasil 31.23% dan k = 3 menghasilkan 26.04%, serta *F1-score* pada k = 5 mendapatkan hasil 26.88% dan k = 3 menghasilkan 21.31%.



Gambar 15 Grafik Perbandingan Hasil Huruf Kapital



Sementara Gambar 16 bagian huruf non kapital dapat disimpulkan bahwa nilai pengujian *accuracy* tertinggi pada huruf kapital yaitu pada $k = 3$ dengan hasil 26.28% sedangkan $k = 5$ dengan hasil 25% dan $k = 7$ dengan hasil 22.44%. Selanjutnya nilai rata-rata *precision* tertinggi yaitu pada $k = 5$ dengan hasil 28.85% sedangkan $k = 3$ dengan nilai 27.27% dan $k = 7$ dengan nilai 22.12%. Berikutnya pada nilai rata-rata *recall* tertinggi yaitu pada $k = 3$ dengan hasil 26.27% sedangkan $k = 5$ dengan nilai 24.61% dan $k = 7$ dengan nilai 23.27%. Dan nilai rata-rata *F1-score* tertinggi yaitu pada $k = 5$ dengan hasil 23.38% sedangkan $k = 3$ dengan nilai 22.7% dan $k = 7$ dengan nilai 20%.



Gambar 16 Grafik Perbandingan Hasil Huruf Non Kapital

4. KESIMPULAN

Berdasarkan pengujian pengenalan tulisan tangan huruf Latin bersambung menggunakan metode KNN dengan ekstraksi ciri Local Binary Pattern yang telah dilakukan mendapatkan kesimpulan berdasarkan hasil performa untuk $k=3$, $k=5$, dan $k=7$. Berdasarkan grafik perbandingan didapatkan hasil terbaik menggunakan nilai $k=7$ dengan tingkat *accuracy* 29% untuk huruf kapital dan nilai $k=3$ dengan tingkat *accuracy* 26%. Kecilnya *accuracy* yang didapatkan dikarenakan *dataset* yang digunakan masih cukup sedikit dan beberapa metode yang digunakan bersifat eksperimen sehingga dirasakan tidak pas saat melakukan proses-proses pada penelitian ini. Penelitian yang telah dilakukan masih sangat bisa dikembangkan kembali dengan penambahan *dataset* yang lebih banyak serta menggunakan metode lain dalam proses ekstraksi ciri ataupun pengenalan sehingga tingkat *accuracy* yang didapatkan juga semakin tinggi.

DAFTAR PUSTAKA

- Al Rivan, M. E., Devella, S., & Saputra, J. (2020). Pengenalan Iris Dengan Normalisasi Menggunakan LBP dan RBF. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 6(2), 122. <https://doi.org/10.24014/coreit.v6i2.9685>
- Andana, A., Widyati, R., & Irzal, M. (2018). Pengenalan Citra Tulisan Tangan Dengan Metode Backpropagation. *Jurnal Matematika Terapan*, 2(1), 36–44.
- Anggraeny, F. T., Munir, M. S., & Purbasari, I. Y. (2020). Histogram Profil Proyeksi sebagai Metode Ekstraksi Fitur pada Pengenalan Karakter Tulisan Tangan. *Prosiding Seminar Nasional Informatika Bela Negara*, 1, 164–168. <https://doi.org/10.33005/santika.v1i0.44>
- Aranta, A., Bimantoro, F., & Putrawan, I. P. T. (2020). Penerapan Algoritma Rule Base dengan Pendekatan Hexadesimal pada Transliterasi Aksara Bima Menjadi Huruf Latin. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTika)*, 2(1), 130–141. <https://doi.org/10.29303/jtika.v2i1.96>
- Bimantoro, F., Aranta, A., Nugraha, G. S., Dwiyanaputra, R., & Husodo, A. Y. (2021). Pengenalan Pola Tulisan Tangan Aksara Bima menggunakan Ciri Tekstur dan KNN. *Journal of Computer Science and Informatics Engineering (J-Cosine)*, 5(1), 60–67.



- <https://doi.org/10.29303/jcosine.v5i1.387>
- Cahyani, S., Wiryasaputra, R., & Gustriansyah, R. (2018). Identifikasi Huruf Kapital Tulisan Tangan Menggunakan Linear Discriminant Analysis dan Euclidean Distance. *Jurnal Sistem Informasi Bisnis*, 8(1), 57. <https://doi.org/10.21456/vol8iss1pp57-67>
- Hidayat, R., & Minati, S. (2019). Comparative Analysis of Text Mining Classification Algorithms for English and Indonesian Qur'an Translation. *IJID (International Journal on Informatics for Development)*, 8(1), 47. <https://doi.org/10.14421/ijid.2019.08108>
- Ilham, F., & Rochmawati, N. (2020). Transliterasi Aksara Jawa Tulisan Tangan ke Tulisan Latin Menggunakan CNN. *Journal of Informatics and Computer Science (JINACS)*, 1(04), 200–208. <https://doi.org/10.26740/jinacs.v1n04.p200-208>
- Maharani, D., Efendi, R., & Johar, A. (2019). Penerapan Augmented Reality Sebagai Media Pembelajaran Pengenalan Aksara Korea (Hangul). *Jurnal Rekursif*, 7(1), 77–90. <https://doi.org/10.33369/rekursif.v7i1.6320>
- Masrani, H., Ruslianto, I., & Ilhamsyah. (2018). Aplikasi Pengenalan Pola Pada Huruf Tulisan Tangan Menggunakan Jaringan Saraf Tiruan Dengan Metode Ekstraksi Fitur Geometri. *Jurnal Coding, Sistem Komputer Untan*, 06(02), 69–78. <https://doi.org/10.26418/coding.v6i2.26674>
- Maulidah, M., Windu Gata, Rizki Aulianita, & Cucu Ika Agustyaningrum. (2020). ALGORITMA KLASIFIKASI DECISION TREE UNTUK REKOMENDASI BUKU BERDASARKAN KATEGORI BUKU. *E-Bisnis: Jurnal Ilmiah Ekonomi Dan Bisnis*, 13(2), 89–96. <https://doi.org/10.51903/e-bisnis.v13i2.251>
- Nikmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2), 421–432. <https://doi.org/10.24176/simet.v10i2.2882>
- Pangestu, R. A., Rahmat, B., & Anggraeny, F. T. (2020). Implementasi Algoritma Cnn Untuk Klasifikasi Citra Lahan Dan Perhitungan Luas. *Jurnal Informatika Dan Sistem Informasi (JIFoSI)*, 1(1), 166–174. <https://doi.org/10.33005/jifosi.v1i1.5>
- Prihatiningsih, S., M, N. S., Andriani, F., & Nugraha, N. (2019). Analisa Performa Pengenalan Tulisan Tangan Angka Berdasarkan Jumlah Iterasi Menggunakan Metode Convolutional Neural Network. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 24(1), 58–66. <https://doi.org/10.35760/tr.2019.v24i1.1934>
- Purbayanti, T. S. (2018). Pengenalan Tulisan Tangan Huruf Latin Dengan Menggunakan Metode K-Nearest Neighbour. *Simki-Techsain*, 02(02), 3–10.

