

# Ekstraksi Topik Pantun di Twitter Menggunakan K-Means Clustering

Nurissaidah Ulinnuha<sup>1, a)</sup> dan Jiphie Gilia Indriyani<sup>1, 2, b)</sup>

<sup>1</sup> Program Studi Matematika UIN Sunan Ampel Surabaya, Indonesia

<sup>2</sup> Program Studi Sastra Indonesia UIN Sunan Ampel Surabaya, Indonesia

<sup>a)</sup> email: nuris.ulinnuha@uinsby.ac.id

<sup>b)</sup> email: jiphiegilia@uinsby.ac.id

## Abstrak

Pantun merupakan bentuk puisi lama di Indonesia. Seiring berkembangnya waktu, eksistensi pantun masa kini dalam masyarakat menjadi keresahan pegiat-pegiat sastra, bahasa dan kebudayaan. Penelitian ini bertujuan menelusuri keberadaan penggunaan pantun di media sosial yang sangat minati anak muda pada era Society 5.0. Terdapat dua jenis pantun yang diteliti yaitu pantun kategori umum dan islami. Metode yang digunakan adalah analisis clustering K-Means untuk menemukan topik pantun yang sering dipakai dalam media sosial. Struktur cluster pantun kategori umum termasuk ke struktur lemah karena nilai koefisien silhouette berada di rentang 0,26-0,5 sedangkan struktur cluster pantun kategori Islami termasuk ke struktur baik hingga kuat karena nilai koefisien silhouetennya berada di rentang 0,4-0,9. Didapatkan hasil bahwa pantun kategori umum masih ramai digunakan di media sosial dengan jenis temanya adalah pantun remaja. Tujuan penggunaan pantunnya lebih dominan ke mengekspresikan perasaan. Sebaliknya pantun kategori islami jarang digunakan di sosial media dengan jenis temanya adalah pantun orang tua. Tujuan penggunaan pantunnya lebih dominan ke syiar agama.

*Kata kunci: pantun, analisis topik, clustering K-Means*

## Abstract

Pantun is an old form of poetry in Indonesia. Over time, the anxiety of the existence of rhymes in society has become the anxiety of literary, language and cultural activists. This study aims to explore the existence of the use of rhymes on social media which was highly encouraged by young people in the era of Society 5.0. There are two types of rhymes studied, namely general and islamic rhymes. The method used is K-Means clustering analysis to find rhyme topics that are often used in social media. The general category of pantun cluster structure belongs to the weak structure because the silhouette coefficient values are in the range of 0.26-0.5 while the Islamic pantun group structure belongs to the good too strong structure because the silhouette coefficient values are in the range of 0.4-0.9. It was found that the general category of rhymes were still widely used on social media with the type of theme being youth rhymes. The purpose of using the rhyme is more dominant to express feelings. On the other hand, the islamic category of rhymes is rarely used on social media with the type of theme being parental rhymes. The purpose of using the rhyme is more dominant to religious symbols.

*Keywords: pantun, topic analysis, K-Means clustering*

## Pendahuluan

Pantun merupakan sebuah bentuk puisi lama di Indonesia[1]. Pantun memiliki bentuk yang khas dan mudah dikenali, selain karena setiap pantun pasti terdiri dari empat larik, pantun juga memiliki rima yang menarik. Indonesia memiliki banyak nama lain dari pantun, setiap daerah menyebut pantun dengan bahasa daerah masing-masing, misal bahasa Jawa menyebut pantun dengan *parikan*, disebut ende-ende dalam bahasa Mandailing, dan disebut *paparegan* dalam Bahasa Madura[2]. Walaupun memiliki sebutan yang beragam, pantun tetap memiliki konsep bahwa larik pertama dan kedua disebut sampiran serta larik ketiga dan keempat disebut isi. Pantun memiliki rima yang menarik karena satu suku kata terakhir yang sama antara larik pertama dengan larik ketiga dan larik kedua dengan larik keempat.

Pantun merupakan objek kajian yang menarik dalam berbagai judul penelitian [3]. Hal ini dikarenakan pantun merupakan sebuah bentuk sastra bersifat umum, dapat digunakan dalam segala acara baik formal seperti kegiatan-kegiatan akademik maupun tidak formal misal dalam acara pernikahan. Berdasarkan jenis penggunaannya, pantun juga dapat dinikmati segala usia, mulai dari pantun nasihat, pantun percintaan, hingga pantun jenaka. Pada beberapa daerah di Indonesia, pantun merupakan bagian dari sebuah seremonial misalnya pada acara pernikahan [3], [4].

Seiring berkembangnya waktu, eksistensi pantun dalam masyarakat menjadi keresahan pegiat-pegiat sastra, bahasa dan kebudayaan. Pantun sebagai produk sastra lama dianggap tren lama oleh generasi muda saat ini. Siti menyampaikan bahwa pantun masih eksis di era teknologi saat ini [5] karena pantun memiliki banyak kelebihan untuk beradaptasi, diantaranya adalah berbentuk sederhana, memuat definisi budaya dan berbudaya dan memuat keindahan dan cita rasa estetika.

Peneliti menelusuri jejak pantun dalam *platform* media sosial lain selain *Facebook*, yaitu pada media sosial *Twitter*. Berdasarkan observasi awal, ditemukan berbagai penelitian dengan objek penelitian pantun dalam berbagai tema. Analisis eksistensi pantun pada artikel ini menggunakan bantuan metode *clustering*, yang merupakan salah satu teknik dalam data mining. *Clustering* adalah seni menemukan kelompok dalam data [6]. Metode *clustering* yang diimplementasikan bertujuan untuk menganalisis *tweet* dengan pengelompokan data secara otomatis. Peneliti mengusulkan metode *clustering* K-Means dalam menganalisis topik pada penggunaan pantun oleh pengguna *Twitter*. Metode ini akan mempartisi data ke dalam *cluster*/grup sehingga data yang mempunyai ciri sama berada pada satu *cluster* sedangkan data yang mempunyai ciri berbeda berada pada *cluster* lain.

Implementasi data mining pada data teks untuk penggalian topik telah banyak dilakukan. Diantaranya penelitian yang dilakukan oleh Bagus dan Gangga dalam memodelkan topik pengguna *Twitter* mengenai aplikasi "Ruangguru" [7]. Diketahui bahwa pengelompokan menghasilkan 28 buah topik dengan topik yang sering diperbincangkan adalah diskon ruangguru. Penelitian lain dilakukan oleh Nugraha [8] yang memanfaatkan metode *clustering* dalam mendeteksi topik pada data skala besar tanpa mereduksi kualitas topik yang didapatkan. Februiyanti dan tim juga melakukan penelitian *clustering* untuk mengetahui kecenderungan mahasiswa terhadap pemberitaan dan mengetahui topik yang seringkali muncul di *Twitter* @unisbank [9]. Pada penelitian terdahulu tersebut, dapat disimpulkan bahwa metode *clustering* telah banyak diimplementasikan dalam deteksi topik di media sosial.

Salah satu metode *clustering* yang populer adalah metode K-Means dikarenakan kemampuan mengelompokkan data dalam jumlah yang besar dengan waktu komputasi yang cepat dan efisien [10]. Yusril dan tim mengimplementasikan *clustering* K-Means pada data *tweets* Gojek Indonesia

[11]. Hasil dari penelitian ini didapatkan sejumlah dua *cluster tweet*. Berdasarkan perhitungan rata-rata jumlah *retweet* di setiap *cluster*, ditemukan bahwa jenis konten yang paling banyak di-*retweet* adalah program yang berkaitan dengan kuis dan pengenalan produk untuk Gojek Indonesia. Diketahui pula nilai *silhouette coefficient* sebesar 0.7 yang berarti metode K-Means memiliki performa yang baik dalam mengelompokkan data.

Penelitian lain yang dilakukan Prasetya mengelompokkan data teks berita dengan membandingkan K-Means dan Single Linkage. Didapatkan kesimpulan bahwa penggunaan data *Twitter* metode K-Means mempunyai indikasi dapat menghasilkan nilai validasi *cluster* yang lebih tinggi daripada Single Linkage [12]. Hal ini dapat dikarenakan penentuan titik *centroid* awal pada data mendekati data yang dominan. Algoritma yang diimplementasikan pada clustering data adalah K-Means. K-Means adalah algoritma berbasis centroid, atau algoritma yang menghitung jarak untuk menetapkan titik ke sebuah cluster. K-Means termasuk salah satu metode yang populer digunakan dalam kategori metode non hirarki dikarenakan kesederhanaan dalam implementasi serta durasi waktu perhitungan komputer yang relatif pendek. K-Means juga memiliki performa yang cukup baik dalam membentuk struktur cluster yang kuat dengan nilai *silhouette coefficient* di atas 0.5, diantaranya dalam penelitian Yuan pada data iris [13], penelitian Battaglia pada data evaluasi penalaran siswa [14] dan Nanjundan pada data sintesis [15].

Berdasarkan penelitian sebelumnya yang telah disebutkan, K-Means memiliki performa yang baik dalam *clustering* data teks. Oleh karena itu, K-Means cocok digunakan pada penelitian ini untuk mengetahui topik pada penggunaan pantun di media sosial *Twitter*. Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada data yang digunakan. Belum ada penelitian yang meneliti penggalian topik pada data *Twitter* yang berkaitan dengan bait pantun. Diharapkan dengan adanya penelitian ini dapat memberikan informasi eksistensi penggunaan pantun sebagai warisan budaya hingga masa kini.

## Metode

Data yang dipakai pada penelitian ini merupakan data primer yang berasal dari dataset *Twitter*. Data tersebut diambil mulai tanggal 20 November 2021 sampai 18 September 2022 selama 37 minggu atau 262 hari. Data *Twitter* yang dikumpulkan diambil dengan larik pantun yang populer digunakan yaitu: "ikan hiu makan tomat", "berakit-rakit ke hulu", "satu titik dua koma" dan "buah manggis buah kedondong". Sedangkan larik yang digunakan untuk mencari pantun islami adalah: "naik unta pergi berhaji", "anak ayam turunnya lima", "kalau bulan rindukan mentari" dan "wajah cantik dibawa ke tanah" .

Penelitian ini dikerjakan dalam tahapan sebagai berikut:

### Tahap 1. Pengumpulan Data

Aplikasi yang digunakan untuk mengambil data pantun umum adalah TAGS yang merupakan template Google Sheet yang dapat mengambil koleksi data *Twitter* secara otomatis dengan memasukkan kata kunci.

### Tahap 2. Praproses Data

Langkah pertama yang dilakukan adalah menghapus username, *link* dan emoji. Selanjutnya, mengubah seluruh huruf menjadi huruf kecil (*case folding*), menentukan larik pertama, kedua, ketiga dan keempat, dan menyaring pantun dengan membuang *tweet* yang tidak mengandung pantun. Setelah itu baru dilanjutkan ke menghapus tanda hubung (*remove punctuations*), memotong teks menjadi kumpulan kata (*tokenizing*) kemudian kemudian menghapus kata yang paling sering muncul (*stopwords removal*) Output akhir dari praproses adalah kumpulan kata yang menjadi nama variabel

yang digunakan pada tahap *clustering*. Nilai dari variabelnya diperoleh dari pembobotan kata di tahap selanjutnya.

#### Tahap 3. Pembobotan Kata.

Proses ini mengubah data teks menjadi data angka yang siap diolah dalam tahap *clustering*. Setiap kata diberikan bobot angka yang dihitung menggunakan Persamaan 1 dan 2.

$$TF-IDF(t,d) = TF_{(t,d)} \times IDF_{(t)} \quad (1)$$

$$IDF(t) = \log \frac{N}{N(t)+1} \quad (2)$$

TF(t,d) menunjukkan frekuensi kemunculan kata t dalam dokumen d, sedangkan N menunjukkan jumlah dokumen, N(t) menunjukkan jumlah dokumen yang mengandung kata t. Adanya penambahan angka satu pada persamaan IDF bertujuan mencegah kondisi penyebut N(t) = 0.

Semakin sering kemunculan kata dalam satu dokumen akan menyebabkan frekuensi kata (TF) yang lebih besar sedangkan semakin jarang kemunculan kata dalam keseluruhan dokumen akan menyebabkan frekuensi dokumen terbalik (IDF) yang lebih besar.

#### Tahap 4. *Clustering* Data

Hasil pembobotan kata berupa nilai TF-IDF masing-masing kata kemudian menjadi nilai input dari variabel *clustering*. Metode *clustering* dapat menyediakan ringkasan tweet yang koheren, artinya informasi yang dihasilkan dari pengelompokan tersebut memiliki keterkaitan yang baik, sehingga dapat memberikan informasi ringkas tentang keseluruhan tweet dengan topik tertentu. *Clustering* adalah tugas membagi populasi menjadi beberapa kelompok berdasarkan kemiripan karakteristik data [16],[17],[18]. Tujuan *clustering* teks pada penelitian ini adalah untuk membentuk kelompok dengan kesamaan topik Twitter. Hasil *clustering* nantinya akan digunakan untuk mengekstraksi topik yang sering muncul dengan larik pantun terkait. Algoritma *clustering* yang digunakan pada penelitian ini adalah K-Means.

Cara kerja algoritma K-Means dijelaskan seperti berikut[16]:

- 1) Tentukan jumlah cluster (k).
- 2) Tetapkan nilai centroid (titik tengah) dengan terlebih dahulu mengacak dataset dan kemudian secara acak pilih k titik data untuk dijadikan centroid.
- 3) Lakukan iterasi sampai tidak ada perubahan pada centroid, yaitu penetapan titik data ke cluster tidak berubah.
- 4) Hitung jarak Euclidean antara titik data dan semua centroid.
- 5) Tetapkan setiap titik data ke cluster dengan jarak terkecil.
- 6) Hitung centroid untuk cluster dengan mengambil rata-rata dari semua titik data yang dimiliki setiap cluster.

Setelah *clustering* dilakukan, selanjutnya masuk ke tahap validasi bertujuan untuk mengevaluasi model *clustering* metode K-Means.

#### Tahap 5. Validasi Cluster

Validasi *cluster* bertujuan untuk mengetahui performa algoritma *clustering* dalam mengelompokkan data. Salah satu validasi *cluster* yang sering digunakan adalah *silhouette coefficient*. *Silhouette coefficient* [11] mengukur seberapa mirip suatu titik dengan klasternya sendiri dibandingkan dengan klaster lainnya. Semakin nilai *silhouette coefficient* mendekati 1, maka semakin baik struktur *cluster*. Persamaan *silhouette coefficient* yaitu:

$$s(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} \quad (3)$$

Variabel a adalah jarak rata-rata suatu titik ke titik lain pada kelompok yang sama sedangkan variabel b adalah nilai minimum jarak dari rata-rata suatu titik ke semua titik pada kelompok lain.

#### Tahap 6. Analisis Topik

Hasil *clustering* kemudian dianalisis dari kemunculan kata yang paling sering muncul di masing-masing kelompok. Hal ini untuk mengetahui topik apa yang sering dikaitkan dengan pantun. Pada pembahasan ini juga dianalisis struktur pantun. Tahap ini menganalisis diksi yang digunakan sebagai pedoman pengklasifikasian jenis pantun berdasarkan tema, yaitu pantun anak, pantun remaja, dan pantun orang tua.

## Hasil dan Diskusi

### 1. Pengumpulan Data

Contoh hasil pengambilan data tweet dapat dilihat pada Tabel 1.

**Tabel 1.** Sampel data twitter untuk larik pantun umum dan Islami

No	Data <i>tweet</i>	Jenis
1	Ikan hiu makan tomat, ayo semangat!!	Umum
2	Ikan hiu makan tomat. I love u so much	Umum
3	Anak ayam turunnya lima Mati seekor tinggal lah empat Kita wajib bela ULAMA Agar lempang jalan ke akhirat	Islami
4	Kalau bulan rindukan mentari Tentu malam akan rindu siang Kalau hati cinta Ilahi Tentu dirinya akan merasa tenang.	Islami

Pantun islami diambil secara manual dikarenakan jarang muncul di sosial media, sementara aplikasi TAGS hanya bisa mengambil data dari 6-9 hari terakhir. Oleh karena itu jumlah datanya tidak terlalu banyak di kisaran 30-60 data.

**Tabel 2.** Jumlah Data Tweet yang Terkumpul

Pantun Umum		Pantun Islami	
Frase	Jumlah	Frase	Jumlah
ikan hiu makan tomat	16.773	naik unta pergi berhaji	42
satu titik dua koma	15.045	anak ayam turunnya lima	30
berakit-rakit ke hulu	9.135	Wajah cantik dibawa ke tanah	36
buah manggis buah kedondong	4.470	Kalau bulan rindukan mentari	52
<b>Total Pantun Umum</b>	<b>45.423</b>	<b>Total Pantun Khusus</b>	<b>160</b>

Tabel 2 menunjukkan jumlah data tweet yang berhasil dikumpulkan. Dapat dilihat bahwa hasil pantun umum sejumlah 45.423 buah, sedangkan pantun islami hanya 160 buah, hasil tersebut berbanding 285:1.

### 2. Praproses Data

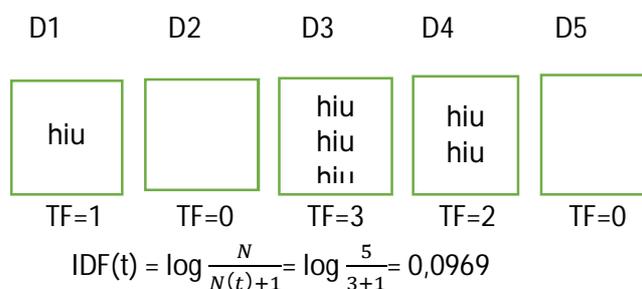
Contoh tahapan praproses akan dipaparkan pada Tabel 3. Input tweet yang digunakan adalah "RT @keijn: @hyeaongjun @myorrow @seeund Ikan hiu makan tomat, i love you, somat <https://t.co/HhA7M8JgSL>"

**Tabel 3.** Praproses Data

No	Tahap Pra Proses	Output
1	Menghapus username, link dan emoji	ikan hiu makan tomat, i love you, somat
2	<i>Case folding</i>	ikan hiu makan tomat, i love you, somat
3	Menyusun larik	ikan hiu makan tomat → larik 1 i love you, somat → larik 2
4	Menyaring pantun (pengecekan larik kedua apakah berakhiran "at" seperti larik pertama.)	ikan hiu makan tomat → larik 1 i love you, somat → larik 2 (Catatan: sudah sesuai, bisa dilanjutkan)
5	<i>Remove punctuations</i>	ikan hiu makan tomat → larik 1 i love you somat → larik 2
6	<i>Tokenizing</i>	ikan   hiu   makan   tomat   i love   you   somat
7	<i>Stopwords removal</i>	ikan   hiu   makan   tomat   i love   you   somat (Catatan: tidak ada yang dihapus karena tidak ada stopword. Contoh stopword: adalah, masih, menurut dst.)

### 3. Pembobotan Kata

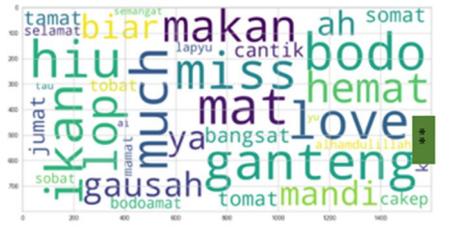
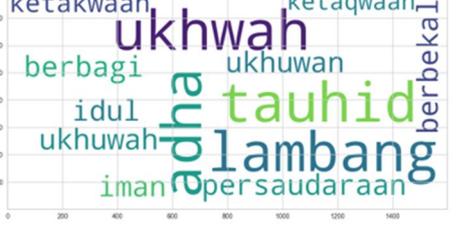
Hasil dari proses pembobotan kata adalah hasil perkalian antara nilai TF dan IDF, yang menghasilkan bobot yang lebih rendah jika kata tersebut sering muncul pada setiap dokumen yang ada di dalam koleksi, sedangkan bobot akan lebih tinggi jika kata tersebut jarang muncul pada setiap dokumen. Ilustrasi perhitungan pembobotan ditunjukkan di Gambar 1.

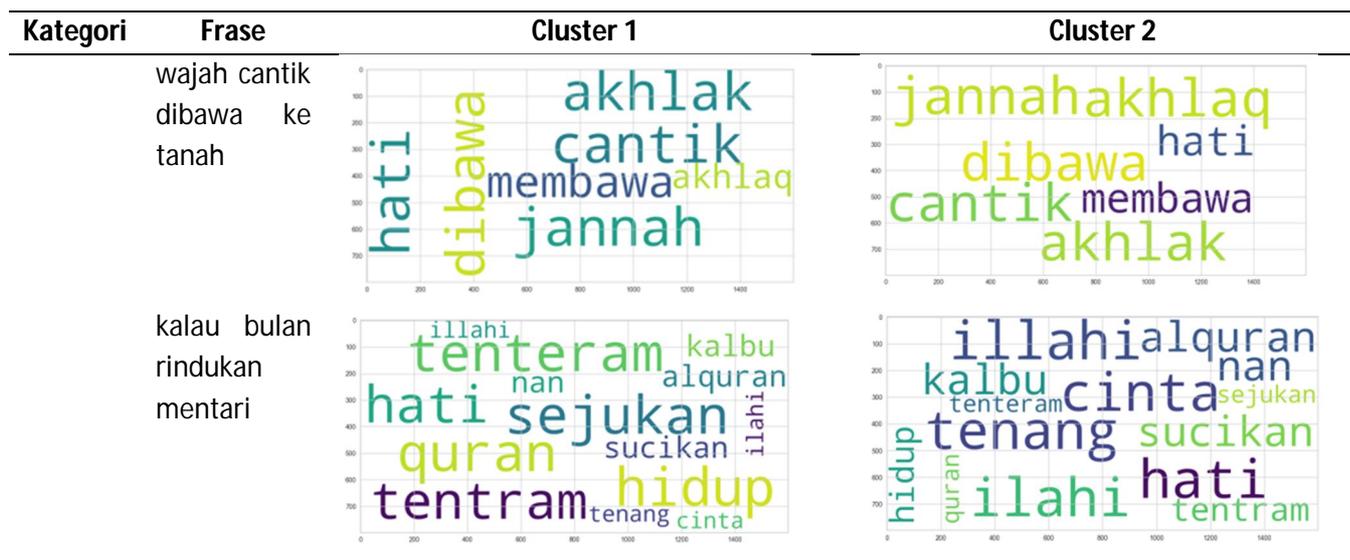
**Gambar 1.** Ilustrasi Pembobotan TF-IDF

### Tahap 4. Clustering Data

Tahap pengelompokan dilakukan berdasarkan data dalam matriks TF-IDF yang telah dikembangkan sebelumnya dengan menggunakan algoritma K-Means. Diambil maksimal 15 term dengan frekuensi tertinggi dari hasil cluster untuk mengetahui topik populer pantun. *Word Cloud* kemudian dibuat untuk menampilkan data teks secara visual agar mudah dipahami dan diidentifikasi topiknya. Hasil wordcloud ditunjukkan oleh Tabel 4.

Tabel 4. Wordcloud setiap cluster pada pantun

Kategori	Frase	Cluster 1	Cluster 2
Umum	ikan hiu makan tomat		
	satu titik dua koma		
	berakit ke hulu		
	buah manggis buah kedondong		
Islami	naik unta pergi berhaji		
	anak ayam turunnya lima		



Tahap 5. Validasi Cluster

Setelah proses *clustering* selesai, dilakukan pengukuran performa hasil *clustering* dengan mengukur rata-rata jarak intra dan antar *cluster* yang dijelaskan pada Tabel 5. Dapat diamati bahwa seluruh struktur *cluster* pantun kategori umum termasuk ke struktur lemah karena berada di rentang 0,26-0,5. Jumlah fitur (*term*) yang banyak dan variatif menyebabkan data menjadi kompleks untuk dikelompokkan. Pada pembahasan penelitian ini dibatasi hanya dua cluster dikarenakan setelah melalui uji coba, semakin besar jumlah cluster, nilai *silhouette coefficient* cenderung menurun. Oleh karena itu, penelitian ini hanya berfokus pada analisis dua cluster saja.

Struktur *cluster* pantun kategori islami termasuk ke struktur lemah untuk frase “naik unta pergi berhaji”, sedangkan tiga pantun islami lainnya masuk struktur kuat dengan rentang nilai *silhouette coefficient* diatas 0.5. Hal ini karena sedikitnya variasi pada isi pantun islami.

Tabel 5. Validasi Hasil Clustering

Kategori	Frase	Silhouette coefficient	Struktur cluster
Umum	ikan hiu makan tomat	0,2768	Lemah
	satu titik dua koma	0,4120	Lemah
	berakit ke hulu	0,3406	Lemah
	buah manggis buah kedondong	0,2576	Lemah
Islami	naik unta pergi berhaji	0,4107	Lemah
	anak ayam turunnya lima	0,7210	Kuat
	wajah cantik dibawa ke tanah	0,7013	Kuat
	kalau bulan rindukan mentari	0,8546	Kuat

Tahap 6. Analisis Topik

Setiap cluster diambil 15 term tertinggi skor bobotnya untuk dianalisis topik pantunnya. Tabel 6 menunjukkan term yang memiliki frekuensi paling tinggi di masing-masing pantun dan clusternya.

Tabel 6. Top terms pada pantun di Twitter

Pantun	Cluster	Top terms
Ikan hiu makan tomat	1	somat lapyu tomat bodo lop love yu mat ai makan so ah tau bodo amat

Pantun	Cluster	Top terms
	2	<i>much so love you miss</i> mat ganteng bodo ikan hiu makan lop hemat gausah mandi
Satu titik dua koma	1	cantik bara makji ayang nya mimin rosa kak cakep wa cantik mempesona ganteng gapake itik
	2	cantik sehun ganteng koma taeyong yauda titik si gue ya rosā nya wahai no
Berakit-rakit ke hulu	1	Bersakit sakit menang melulu rebahan kerjaannya bersenang senang enhypen vote daesang mm sakit bersakit nyusul status penghulu
	2	penghulu mantan kesepian ayo biar mari sakit beda kepastian jalan love you kerjaannya jadian temenan
Buah manggis buah kedondong	1	manis hai follow bayu hy manies cubit pipinya rj dikit ayo army vte kenalan lirik
	2	cewe kenalan manis hy bayu beli ayo manies vte sims ayank the army mnies donat
Naik unta pergi berhaji	1	tauhid ukhuwah adalah lambang persaudaraan idul berbagi berbekal iman ukhuwah ukhuan ketakwaan ketaqwaan
	2	ketaqwaan iman berbagi idul berbekal persaudaraan lambang adha ukhuwah ukhuan ketakwaan tauhid ukhwah
Anak ayam turunnya lima	1	hidup beragama mesti sesat jalan mengikut makrifat alim ulama
	2	jalan ulama alim mengikut makrifat hidup sesat mesti beragama
Wajah cantik dibawa ke tanah	1	cantik hati jannah dibawa akhlak membawa akhlaq
	2	dibawa akhlak jannah akhlaq cantik membawa hati
Kalau bulan rindukan mentari	1	sejukan quran hidup tenteram tentram hati alquran sucikan nan kalbu tenang illahi ilahi cinta
	2	ilahi cinta tenang hati illahi alquran sucikan nan kalbu tentram hidup sejukan quran tenteram

Hasil analisis data memperoleh bahwa pantun yang lariknya dengan larik: "ikan hiu makan tomat", "satu titik dua koma", "berakit-rakit ke hulu", dan "buah manggis buah kedondong" memiliki tujuan penggunaan yang beragam. Dapat disimpulkan bahwa jenis pantun berdasarkan isinya yang sering muncul adalah pantun cinta karena pantun-pantun tersebut digunakan oleh pengguna media sosial untuk mengekspresikan perasaan seperti cinta, kagum, dan kesal. Ini berarti bahwa isi atau tema dari pantun-pantun ini berkaitan dengan perasaan romantis atau emosional. Contoh: "ikan hiu makan tomat, i love you so much" dan "berakit-rakit ke hulu berenang-renang ke tepian, mari kita ke penghulu supaya ada kepastian".

Selain itu, pantun juga ramai digunakan fans K-pop untuk mendukung idola serta berinteraksi dengan sesama fans dalam media sosial. Contoh: "satu titik dua koma, taeyong ganteng siapa yang punya", "buah manggis buah kedondong, army yang manis ayo vote dong", dan "berakit-rakit dahulu, berenang-renang ke tepian, bersakit-sakit vote mm dahulu, enhypen menang daesang kemudian". Di sisi lain, pantun juga digunakan mempromosikan barang dagangan oleh pengguna

sosial media untuk menarik perhatian calon pembeli, seperti: “buah manggis buang kedondong, hai manis beli the sims di aku dong”.

Selanjutnya, berdasarkan pengguna, jenis pantun yang dipakai adalah pantun remaja yang merujuk pada kehidupan individu dalam kelompok usia yang lebih muda. Pantun yang sering digunakan oleh remaja umumnya mengandung makna terkait dengan topik seperti asmara, yang ramai digunakan pada pantun ikan hiu makan tomat, kemudian topik perkenalan seperti pantun “buah manggis buah kedondong, hai manis kenalan dong” dan topik perasaan, seperti “satu titik dua koma, kamu cantik siapa yang punya” yang menyampaikan perasaan kekaguman.

Lain hal dengan pantun islami, berdasarkan hasil analisis data, tujuannya hanya terbagi dua. Pertama, pantun digunakan untuk merayakan hari raya. Pantun yang memuat tujuan ini adalah larik “naik unta pergi berhaji” untuk menyampaikan ucapan selamat Idul Adha. Kedua, pantun digunakan untuk syiar menyampaikan nasihat dan pesan moral. Nasihat tersebut diantaranya bertujuan untuk mengajak mengikuti jalan alim ulama, menjaga hati dan akhlak serta membaca kitab suci Al-Qur’an agar hidup menjadi lebih tenteram. Pantun yang menyertakan tujuan ini adalah larik: “anak ayam turunnya lima”, “wajah cantik dibawa ke tanah”, dan “kalau bulan rindukan mentari”. Dapat disimpulkan bahwa jenis pantun islami berdasarkan isinya yang sering muncul adalah pantun nasihat atau bisa juga disebut sebagai pantun agama sedangkan jenis pantun islami berdasarkan temanya adalah pantun orang tua. Disebut pantun orang tua karena pantun Islami seringkali mengandung pesan moral dan nasihat yang bersifat bijak, yang sering diasosiasikan dengan orang tua yang memberikan panduan dan nasihat kepada generasi muda, dalam konteks budaya dan agama Islam.

## Kesimpulan

Banyaknya penggunaan pantun di twitter menunjukkan bahwa eksistensi pantun yang semula berupa sastra lisan telah beralih ke sastra tulis di media sosial. Eksistensi pantun di kategori umum sangat tinggi dengan jumlah tweet mencapai puluhan ribu. Tujuan penggunaannya beragam mulai dari pengekspresian diri, mendukung idola hingga sarana promosi barang dagangan. Sebaliknya, eksistensi pantun di kategori islami tidak begitu banyak, hanya mencapai ratusan. Tujuan penggunaannya adalah syiar agama juga untuk memperingati hari raya. Dari jumlah larik, jenis pantun yang sering digunakan oleh masyarakat Twitter adalah pantun dua baris (karmina) untuk kategori umum; dan pantun empat baris untuk kategori islami. Dari tema, jenis pantun yang lebih sering digunakan adalah remaja, untuk kategori umum; dan orangtua untuk kategori islami. Dari isi, jenis pantun yang sering digunakan adalah pantun cinta, pantun peribahasa dan pantun dagang untuk kategori umum; dan pantun agama untuk kategori islami. Dapat dinyatakan bahwa pantun yang merupakan bentuk sastra lisan didapati masih eksis di masa kini. Eksistensi pantun mengalami perubahan bentuk dari sastra lisan menjadi sastra tulis yang banyak digunakan di media sosial Twitter.

## Referensi

- [1] D. E. Maulina, “Keanekaragaman Pantun,” *Semantik*, vol. 1 No 1, no. 1, pp. 107–121, 2012, [Online]. Available: <http://e-journal.stkipsiliwangi.ac.id/index.php/semantik/article/view/103>
- [2] M. S. Mahayana, “Pantun Sebagai Potret Sosial Budaya Tempatan Perbandingan Pantun Melayu, Jawa, Madura, Dan Betawi,” *Jurnal Kritik*, vol. 4, pp. 85–100, 2013.
- [3] A. Aslan and A. Yunaldi, “Budaya Berbalas Pantun Sebagai Media Penyampaian Pesan Perkawinan Dalam Acara Adat Istiadat Perkawinan Melayu Sambas,” *Transformatif*, vol. 2, no. 2, pp. 111–122, 2019, doi: 10.23971/TF.v2i2.962.

- [4] A. Yuliansyah, "Struktur Dan Fungsi Pantun Dalam Upacara Adat Perkawinan Melayu Tanjung Hulu Kecamatan Pontianak Timur," *Jurnal Pendidikan Bahasa*, vol. 8, no. 2, p. 281, 2019, doi: 10.31571/bahasa.v8i2.1432.
- [5] K. Pantun, R. dan Leluhur Bangsa Dulu, and K. dan Selamanya SITI HAJAR CHE MAN ABSTRAK, "Siti Hajar Che Man 75," *International Journal of the Malay World and Civilisation (Iman)*, vol. 1, no. 1, p. 2013.
- [6] J. E. Gentle, L. Kaufman, and P. J. Rousseuw, *Finding Groups in Data: An Introduction to Cluster Analysis.*, vol. 47, no. 2. New Jersey: Wiley-Interscience, 1991. doi: 10.2307/2532178.
- [7] B. W. Arianto and G. Anuraga, "Topic Modeling for Twitter Users Regarding the 'Ruangguru' Application," *Jurnal ILMU DASAR*, 2020. <https://jurnal.unej.ac.id/index.php/JID/article/view/17112/8412> (accessed Dec. 06, 2021).
- [8] Christopher Nugraha, "Pendeteksian topik pada twitter menggunakan online eigenspace-based fuzzy c-means clustering untuk big data = Topic detection on twitter using online eigenspace-based fuzzy c-means clustering for big data / Christhoper Nugraha." 2019. Accessed: Dec. 06, 2021. [Online]. Available: <http://lontar.ui.ac.id>
- [9] H. Februariyanti, J. S. Wibowo, D. B. Santoso, and M. Sukur, "Analisis Kecenderungan Informasi Menggunakan Algoritma Hierarchical Agglomerative Clustering," *INFORMATIKA*, vol. 13, no. 1, p. 9, Jun. 2021, doi: 10.36723/juri.v13i1.247.
- [10] B. A. Dwiarni and B. Setiyono, "Akuisisi dan Clustering Data Sosial Media Menggunakan Algoritma K-Means sebagai Dasar untuk Mengetahui Profil Pengguna," *Jurnal Sains dan Seni*, vol. 8, no. 2, pp. 2337–3520, 2019, [Online]. Available: <https://apps.twitter.com/>
- [11] A. N. Yusril, I. Larasati, and Q. Aini, "Implementasi Text Mining Untuk Advertising Dengan Menggunakan Metode K-Means Clustering Pada Data Tweets Gojek Indonesia," *Sistemasi*, vol. 9, no. 3, p. 586, Sep. 2020, doi: 10.32520/stmsi.v9i3.924.
- [12] P. Kurniawan, E. Utami, and A. Sunyoto, "Perbandingan Teknik Pengklasteran Dalam Visualisasi Data Teks Bahasa Indonesia," 2017.
- [13] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J (Basel)*, vol. 2, no. 2, pp. 226–235, 2019, doi: 10.3390/j2020016.
- [14] O. R. Battaglia, B. di Paola, and C. Fazio, "K-means clustering to study how student reasoning lines can be modified by a learning activity based on Feynman's unifying approach," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 6, pp. 2005–2038, 2017, doi: 10.12973/eurasia.2017.01211a.
- [15] S. Nanjundan, S. Sankaran, C. R. Arjun, and G. P. Anand, "Identifying the number of clusters for K-Means: A hypersphere density based approach," 2019. [Online]. Available: <http://arxiv.org/abs/1912.00643>
- [16] N. Ulinuha and S. A. Sholihah, "Analisis Cluster untuk Pemetaan Data Kasus Covid-19 di Indonesia Menggunakan K-Means," *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, vol. 9, no. 2, pp. 27–31, 2021.
- [17] E. Syafaqoh, dkk, "Klasterisasi Penyandang Masalah Kesejahteraan Sosial (PMKS) Di Kabupaten Bojonegoro Menggunakan Algoritma K-Medoids," *KUBIK: Jurnal Publikasi Ilmiah Matematika*, vol. 7, no. 2, pp. 78-87, 2022, doi: 10.15575/kubik.v7i2.21653.
- [18] I. Nurandini and A.F. Huda, "Klastering Dokumen dengan Menambahkan Metadata Menggunakan Algoritma COATES," *KUBIK: Jurnal Publikasi Ilmiah Matematika*, vol. 2, no. 2, pp. 39-44, 2017, doi: 10.15575/kubik.v2i2.1859.