

Mengindetifikasi Kemiripan Kalimat Bahasa Banjar Menggunakan Metode Cosine Similarity Berbasis Android

Delia Virna Putri Arindi¹, Ahmad Bagus Setiawan², Ratih Kumala Sari³

^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Nusantara PGRI Kediri

E-mail: ¹arindiputri13@gmail.com, ²ahmadbagus@unpkediri.ac.id, ³ratih.workmail@gmail.com

Abstrak – Bahasa adalah system lambang atau simbol bunyi yang berkembang berdasarkan suatu aturan yang disepakati oleh pemakaiannya. Setiap lambang bahasa memiliki makna atau konsep. Karena setiap lambing bunyi memiliki atau menyatakan suatu konsep atau makna. Apalagi dengan bahasa daerah, bahasa daerah biasanya di gunakan menjadi bahasa sehari-hari di daerah tersebut. Dengan menggunakan bahasa daerah yang baik dan benar itu menunjukkan bahwa kita menghargai bahasa yang ada di Indonesia. Dialek Banjar Kuala umumnya dipakai oleh penduduk asli sekitar kota Banjarmasin, Martapura, dan Pelaihari, sedangkan dialek Banjar Hulu adalah bahasa Banjar yang dipakai penduduk daerah hulu sungai umumnya yaitu daerah kabupaten Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara (dan Balangan) serta Tabalong. Cosine Similarity adalah ukuran kesamaan antara dua buah vektor dalam sebuah ruang dimensi yang didapat dari nilai cosinus sudut dari perkalian dua buah vektor yang dibandingkan karena cosinus dari 00 adalah 1 dan kurang dari 1 untuk nilai sudut yang lain, maka nilai similarity dari dua buah vektor dikatakan mirip ketika nilai dari cosine similarity adalah 1.

Kata Kunci — Bahasa, Dialek, Cosine Similarity

1. PENDAHULUAN

Bahasa adalah *system* lambang atau simbol bunyi yang berkembang berdasarkan suatu aturan yang disepakati oleh pemakaiannya. Setiap lambang bahasa memiliki makna atau konsep. Karena setiap lambing bunyi memiliki atau menyatakan suatu konsep atau makna. Apalagi dengan bahasa daerah, bahasa daerah biasanya di gunakan menjadi bahasa sehari-hari di daerah tersebut. Dengan menggunakan bahasa daerah yang baik dan benar itu menunjukkan bahwa kita menghargai bahasa yang ada di Indonesia. Fungsi bahasa adalah sebagai wahana komunikasi bagi manusia, baik komunikasi lisan maupun komunikasi tulis. Bahasa tidak dapat dilepaskan dari kegiatan hidup masyarakat, yang di dalamnya sebenarnya terdapat status dan nilai-nilai social. Bahasa selalu mengikuti warna warni di kehidupan manusia. Apalagi dengan bahasa daerah yang lebih lekat dengan kebudayaan yang ada di masing-masing daerah [3].

Orang Melayu yang tinggal di Provinsi Kalimantan Selatan dan di beberapa tempat di Provinsi Kalimantan Timur dan Kalimantan Tengah disebut orang Banjar. Bahasa mereka disebut bahasa Banjar. Informasi tentang orang Banjar lebih banyak jika dibandingkan dengan informasi tentang orang Dayak.

Hapip dalam Suryadikara (1984) memberikan informasi tentang dua dialek bahasa Banjar, yaitu Banjar Kuala dan Banjar Hulu. Adanya dua dialek itu didasarkan pada perbedaan pemakaian kosakata tertentu dan sistem bunyi vokal kedua dialek tersebut. Pada tahun 1977, Djantera Kawi pernah

menulis “Geografi Dialek Kotamadya Banjarmasin” dan ia memberikan informasi tentang perpindahan penduduk yang terjadi sejak tahun 1526 hingga sekarang dari hulu sungai ke Banjarmasin. Dengan demikian, di Kotamadya Banjarmasin terdapat dua dialek bahasa Banjar, yaitu dialek Banjar Kuala dan dialek Banjar Hulu. Dialek Banjar Kuala umumnya dipakai oleh penduduk asli sekitar kota Banjarmasin, Martapura, dan Pelaihari, sedangkan dialek Banjar Hulu adalah bahasa Banjar yang dipakai penduduk daerah hulu sungai umumnya yaitu daerah kabupaten Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara (dan Balangan) serta Tabalong. Penutur dialek Banjar Hulu ini jauh lebih luas dan masih menunjukkan beberapa variasi subdialek lagi yang oleh Den Hamer disebut dengan istilah dialek lokal yaitu seperti Amuntai, Alabiu, Kalua, Kandangan, dan Tanjung. Bahkan Den Hamer cenderung berpendapat bahwa yang dipakai oleh *orang Bukit* yaitu penduduk pedalaman penguasaan Meratus merupakan salah satu subdialek Banjar Hulu pula [5].

Android merupakan sistem operasi mobile. Android tidak membedakan antara aplikasi inti dengan aplikasi pihak ketiga. Application Programming Interface (API) yang disediakan menawarkan akses ke hardware, maupun data data ponsel sekalipun, atau data sistem sendiri. [4]

Cosine Similarity adalah ukuran kesamaan antara dua buah vektor dalam sebuah ruang

dimensi yang didapat dari nilai *cosinus* sudut dari perkalian dua buah vektor yang dibandingkan karena *cosinus* dari 00 adalah 1 dan kurang dari 1 untuk nilai sudut yang lain, maka nilai *similarity* dari dua buah vektor dikatakan mirip ketika nilai dari *cosine similarity* adalah 1. *Cosine similarity* digunakan dalam ruang positif, dimana hasilnya dibatasi antara nilai 0 dan 1. Kalau nilainya 0 maka dokumen tersebut dikatakan mirip jika hasilnya 1 maka nilai tersebut dikatakan tidak mirip Perhatikan bahwa batas ini berlaku untuk sejumlah dimensi, dan *Cosine similarity* ini paling sering digunakan dalam ruang positif dimensi tinggi. Misalnya, dalam Information Retrieval, masing-masing kata/istilah (term) diasumsikan sebagai dimensi yang berbeda dan dokumen ditandai dengan vector dimana nilai masing-masing dimensi sesuai dengan berapa istilah muncul dalam dokumen [1]

2. METODE PENELITIAN

Cosine Similarity adalah ukuran kesamaan antara dua buah vektor dalam sebuah ruang dimensi yang didapat dari nilai *cosinus* sudut dari perkalian dua buah vektor yang dibandingkan karena *cosinus* dari 00 adalah 1 dan kurang dari 1 untuk nilai sudut yang lain, maka nilai *similarity* dari dua buah vektor dikatakan mirip ketika nilai dari *cosine similarity* adalah 1, *Cosine similarity* digunakan dalam ruang positif, dimana hasilnya dibatasi antara nilai 0 dan 1. Kalau nilainya 0 maka dokumen tersebut dikatakan mirip jika hasilnya 1 maka nilai tersebut dikatakan tidak mirip Perhatikan bahwa batas ini berlaku untuk sejumlah dimensi, dan *Cosine similarity* ini paling sering digunakan dalam ruang positif dimensi tinggi. Misalnya, dalam Information Retrieval, masing-masing kata/istilah (term) diasumsikan sebagai dimensi yang berbeda dan dokumen ditandai dengan vector dimana nilai masing-masing dimensi sesuai dengan berapa istilah muncul dalam dokumen [1]

2.1 Data kata bahasa banjar

Tabel 1. Kata Dasar Bahasa Banjar

Banjar Muara	Banjar Banua Anam	Indonesia
Baduhara	bakurinah	dengan sengaja
Bibit	jumpu/ambil	Ambil
bungas/langkar	mulik/baik rupa	Cantik
Caram	Calap	tergenang air
Canggar	Kajung	tegang/ereksi
Ampah	Mara	Arah
Hangkuy	nyaring	Nyaring
Hagan	Gasam	Untuk
gani'i	dangani	Temani
ma-hurup	ma-nukar	mem-beli
padu/padangan	Dapur	Dapur
Hingkat	Kawa	dapat/bisa
Pawa	Wadah	Tempat

himpit/tawak/tukun/hantup	hamput	sambit (lempar)
Aray	himung	Senang
Tiring	Lihat	Memandang
Tingaw	Lihat	Toleh
Balalah	bakunjang	Bepergian
Linger	Tuang	Tuang
Tuti	tadih/hintadi	Tadi
ba-ugah	ba-jauh	men-jauh
Macal	Nakal	Nakal
Balay	langgar	Surau
Tutuy	Catuk	pukul dengan palu
Kaday	warung	Warung
kau'	Nyawa	Kamu
Diaku	Unda	Aku
di sia	di sini	di sini
bat-ku	ampun-ku	punya-ku
ba-cakut	ba-kalahi	Berkelahi
ba-cakut	ba-pingkut	berpegangan pada sesuatu
Diang	Galuh	panggilan perempuan anak
nini laki	Kai	Kakek
Utuh	nanang	panggilan anak lelaki
Uma	Mama	Ibu
Puga	Anyar	Baru
Salukut	Bakar	Bakar
kasalukutan/ka mandahan	kagusangan	Kebakaran
Tajua	Ampih	Berhenti
acil laki	Amang	Paman

Sumber : Balai Bahasa Banjarmasin. 2008. *Kamus Bahasa Banjar Dialek Hulu-Indonesia*. Banjarbaru

2.2 Rumus Cosine Similarity

Berikut ini adalah Rumus dari *Cosine Similarity*

$$\frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Keterangan:

A = vektor

B = vektor

A_i = bobot term i dalam blok A_i

B_i = bobot term i dalam blok B_i

i = jumlah term dalam kalimat

n = jumlah vector

Dimana A merupakan bobot setiap ciri pada vector A, dan B merupakan bobot setiap ciri pada B. jika dikaitkan dengan information retrieval maka A adalah bobot setiap istilah pada dokumen A, dan B merupakan bobot setiap istilah pada dokumen B. Pada penelitian ini digunakan cosine similarity karena citra merupakan salah satu data yang memiliki dimensi tinggi. Pada citra dapat dikatakan bahwa setiap pixel merupakan dimensi yang berbeda dan nilai warna pada setiap pixel tersebut merupakan nilai dari setiap dimensi. Pengukuran kemiripan dapat dilakukan dengan membandingkan dokumen 1 dengan dokumen 2 kemudian sistem akan

menghitung nilai kemiripan. Ai.Bi adalah nilai yang diperoleh dari term A dan term B kemudian kedua nilai tersebut dijumlahkan, kemudian nilai Ai2 semua nilai term dokumen A semua nilainya dipangkatkan dua, begitu juga dengan term Bi2 semua nilai yang diperoleh dipangkatkan dua kemudian semua nilai yang diperoleh dijumlahkan. Dengan memasukkan 2 kalimat yang berbeda kata lalu di perhitungkan menggunakan metode tersebut. Sebagaimana algoritma dari *Cosine Similarity* tersebut .(Ariantini 2016)[1]

2.3 Simulasi Perhitungan

Kalimat A = Aku lawan ikam kada sama lawan aku

Kalimat B = Ikam lawan aku bujur kada sama lawan ikam

Penyelesaian:

Terdapat dua kalimat yang berbeda dengan arti yang sama , yaitu vector A dan vector B, untuk melihat berapa nilai yang di peroleh dari dua kalimat tersebut, maka di buat kan table seperti di bawah ini:

	A	B
Aku	2	1
Ikam	1	2
Lawan	2	2
Kada	1	1
Bujur	0	1
Sama	1	1

A: (2,1,2,1,0,1)

B: (1,2,2,1,1,1)

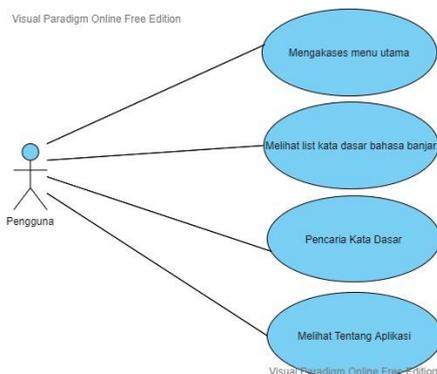
Similarity:

$$= \frac{(2 \times 1) + (1 \times 2) + (2 \times 2) + (1 \times 1) + (0 \times 1) + (1 \times 1)}{\sqrt{2^2 + 1^2 + 2^2 + 1^2 + 0^2 + 1^2} \times \sqrt{1^2 + 2^2 + 2^2 + 1^2 + 1^2 + 1^2}}$$

$$= \frac{10}{4+1+4+1+0+1 \times 1+4+4+1+1+1}$$

$$= \frac{10}{11 \times 12} = \frac{10}{132} = 7,58\%$$

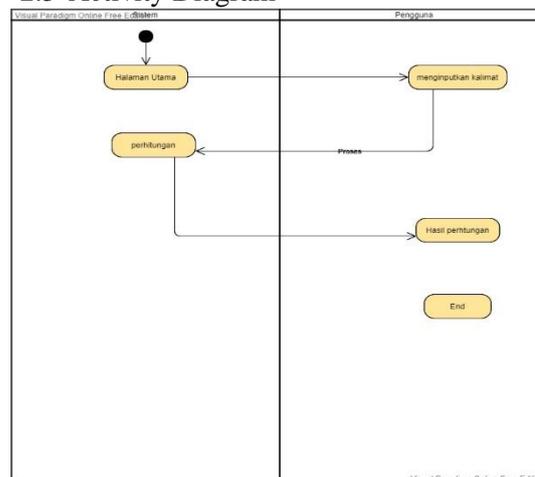
2.4 Use Case Diagram



Gambar 1. Use Case Diagram

Di use case tersebut dijelaskan, ada user yang bisa mengakses menu utama yang berisikan cari kata dan penerjemahannya dan user bisa menginputkan kalimat yang akan di hitung kemiripannya antar kedua teks. Lalu di admin yang mengelola semua dari menu utama sampai hasil kemiripan antara kedua teks tersebut.

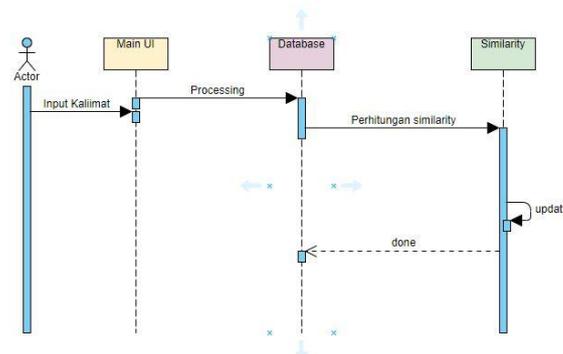
2.5 Activity Diagram



Gambar 2. Activity Diagram

Di sistem membuka halaman utama yang berisikan menu utama, terjemahan dan perhitungan, setelah itu user menginputkan kalimat dan selanjutnya oleh system akan di proses untuk proses perhitungan . setelah itu muncul hasil berapa persen kemiripan antara 2 kalimat tersebut.

2.6 Sequence Diagram



Gambar 3. Sequence Diagram

Pada sequence diagram ini membandingkan antara 2 kalimat yang diinputkan oleh pengguna. Setelah itu akan dihitung menggunakan Similarity berapa persen kemiripan antar 2 kalimat tersebut. Setelah dihitung pengguna akan tahu persentase kemiripan 2 kalimat tersebut

