

TESIS

**STEMMING BAHASA SASAK MENGGUNAKAN KOMBINASI NAZIEF-
ADRIANI DAN PORTER STEMMER**



Disusun oleh:

Nama : Yulita Fatma Andriani
NIM : 17.52.1030
Konsentrasi : Business Intelligence

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2020

TESIS

**STEMMING BAHASA SASAK MENGGUNAKAN KOMBINASI NAZIEF-
ADRIANI DAN PORTER STEMMER**

**SASAK LANGUAGE STEMMING USING COMBINATION OF NAZIEF-
ADRIANI AND PORTER STEMMER**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Yulita Fatma Andriani
NIM : 17.52.1030
Konsentrasi : Business Intelligence

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2020

HALAMAN PENGESAHAN

STEMMING BAHASA SASAK MENGGUNAKAN KOMBINASI NAZIEF-
ADRIANI DAN PORTER STEMMER

SASAK LANGUAGE STEMMING USING COMBINATION OF NAZIEF-
ADRIANI AND PORTER STEMMER

Dipersiapkan dan Disusun oleh

Yulita Fatma Andriani

17.52.1030

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Rabu, 8 Januari 2020

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer



HALAMAN PERSETUJUAN
**STEMMING BAHASA SASAK MENGGUNAKAN KOMBINASI NAZIEF-
ADRIANI DAN PORTER STEMMER**

**SASAK LANGUAGE STEMMING USING COMBINATION OF NAZIEF-
ADRIANI AND PORTER STEMMER**

Dipersiapkan dan Disusun oleh

Yulita Fatma Andriani


17.52.1030

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Rabu, 8 Januari 2020

Pembimbing Utama



Prof. Dr. Ema Utami, S.Si., M.Kom
NIK. 190302037

Anggota Tim Penguji


Dr. Kusriani, M.Kom
NIK. 190302016

Pembimbing Pendamping


Dr. Suwanto Raharjo, S.Si., M.Kom
NIK. 999106


Dr. Wing Wahyu Winarno, MAFIS, Ak.
NIK. 555195


Prof. Dr. Ema Utami, S.Si., M.Kom
NIK. 190302037

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 8 Januari 2020
Direktur Program Pascasarjana

Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Yulita Fatma Andriani
NIM : 17.52.1030
Konsentrasi : Business Intelligence

Menyatakan bahwa Tesis dengan judul berikut:
Stemming Bahasa Sasak Menggunakan Kombinasi Nazief-Adriani Dan Porter Stemmer

Dosen Pembimbing Utama : Prof. Dr. Ema Utami, S.Si., M.Kom.
Dosen Pembimbing Pendamping : Dr. Suwanto Raharjo, S.Si., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 8 Januari 2020
Yang Menyatakan,



UNIVERSITAS AMIKOM
YOGYAKARTA
100723442
000
UNIVERSITAS AMIKOM

Yulita Fatma Andriani

HALAMAN PERSEMBAHAN

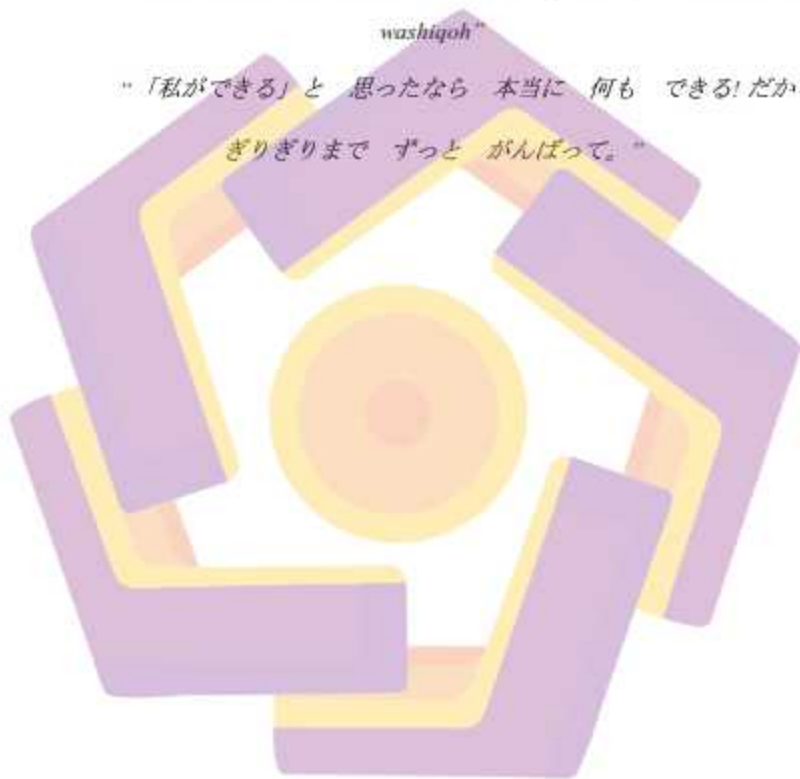
Ucapan syukur saya sampaikan atas kehadiran Allah SWT atas segala nikmat karunia yang telah diberikan kepada saya sehingga saya dapat menyelesaikan thesis saya yang berjudul “Stemming Bahasa Sasak Dengan Kombinasi Nazief-Adriani dan Porter Stemmer”. Shalawat dan salam juga tak lupa saya lantunkan kepada Rasulullah Muhammad SAW yang menjadi pedoman saya untuk menjalani hidup di dunia. Adapun karya ini saya persembahkan untuk:

1. Mamak dan Bapak yang tak pernah putus mengalirkan dukungan, motivasi dan do'a untuk saya.
2. Kakak dan adik perempuan saya yang juga terus memotivasi berserta seluruh anggota keluarga besar saya.
3. Dosen pembimbing utama Ibu Prof Dr. Ema Utami, S.Si, M.Kom dan dosen pembimbing pendamping Bapak Dr. Suwanto Raharjo, S.Si., M.Kom yang membimbing, mengarahkan, dan memperjuangkan saya sehingga dapat menyelesaikan penelitian ini.
4. *My dearest friends who always give me motivation to go on eventhought you guys are not here.*
5. Rekan-rekan kerja yang juga telah memberikan motivasi dan selalu mengerti.
6. Teman-teman sekelas dan seperjuangan, MTI angkatan 19B

HALAMAN MOTTO

*"Al ilmu shoidun wal kitabatu qoiduhu, qoyyid shuyudaka bil hibalil
washiqoh"*

“「私ができる」と思ったなら 本当に 何も できる! だから
ぎりぎりまで ずっと がんばって。”



KATA PENGANTAR

Assalamu 'alaikum wr. wb.

Puji dan syukur penulis persembahkan untuk Allah SWT yang telah memberikan rahmat, hidayah dan kekuatan sehingga penulis dapat menyelesaikan skripsi ini. Tidak lupa *sholawat* serta salam penulis haturkan pada junjungan umat yaitu Nabi Muhammad SAW, yang telah menyebarkan agama Islam sehingga penulis dan seluruh umat Islam dapat merasakan indahnya Islam.

Tesis ini disusun sebagai salah satu syarat kelulusan bagi setiap mahasiswa Universitas AMIKOM Yogyakarta. Selain itu juga merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang program Strata-2 dan untuk memperoleh gelar Magister Teknik Informatika.

Dengan selesainya tesis ini, maka penulis tidak lupa mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. Selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Dr. Kusriani, M.Kom selaku Direktur Program Pascasarjana Universitas AMIKOM Yogyakarta.
3. Ibu Prof Dr. Ema Utami, S.Si, M.Kom selaku pembimbing utama dan Bapak Dr. Suwanto Raharjo, S.Si., M.Kom selaku pembimbing pendamping yang membimbing, mengarahkan, dan memperjuangkan saya sehingga dapat menyelesaikan penelitian ini.

4. Mamak dan Bapak yang selalu memberikan dukungan dan do'a.
5. Dosen penguji seminar proposal tesis, seminar hasil tesis, hingga ujian tesis sehingga penelitian ini dapat terselesaikan dengan baik.
6. Bapak dan Ibu Dosen Universitas AMIKOM Yogyakarta yang telah banyak memberikan ilmunya selama penulis kuliah.
7. Teman-teman semasa kuliah.
8. Dinas Perpustakaan dan Kearsipan Provinsi Nusa Tenggara Barat yang telah memberikan banyak referensi untuk penelitian ini.

Yogyakarta, 8 Januari 2020

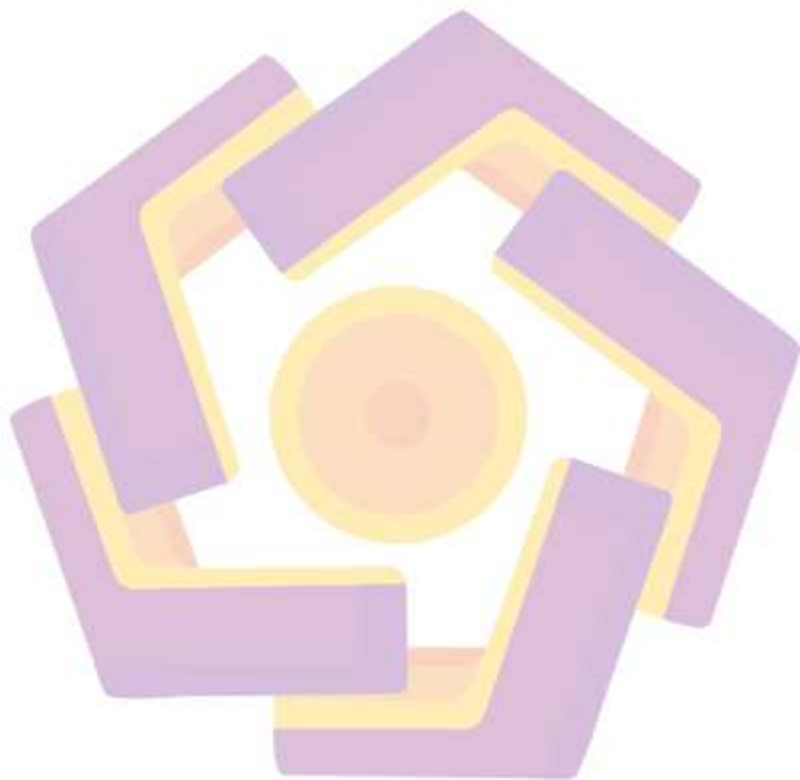
Penulis

DAFTAR ISI

| | |
|---|------|
| HALAMAN JUDUL | ii |
| HALAMAN PENGESAHAN | iii |
| HALAMAN PERSETUJUAN | iv |
| HALAMAN PERNYATAAN KEASLIAN TESIS | v |
| HALAMAN PERSEMBAHAN | vi |
| HALAMAN MOTTO | vii |
| KATA PENGANTAR | viii |
| DAFTAR ISI | x |
| DAFTAR TABEL | xiii |
| DAFTAR GAMBAR | xiv |
| Intisari | xv |
| <i>ABSTRACT</i> | xvi |
| BAB I PENDAHULUAN | 1 |
| 1.1. Latar Belakang Masalah | 1 |
| 1.2. Rumusan Masalah | 3 |
| 1.3. Batasan Masalah | 4 |
| 1.4. Tujuan Penelitian | 4 |
| 1.5. Manfaat Penelitian | 5 |
| BAB II TINJAUAN PUSTAKA | 6 |
| 2.1. Tinjauan Pustaka | 6 |
| 2.2. Keaslian Penelitian | 9 |
| 2.3. Landasan Teori | 14 |

| | |
|---|-----------|
| 2.3.1. Bahasa Sasak | 14 |
| 2.3.2. Tingkat Tutur Bahasa Sasak | 15 |
| 2.3.3. Dialek Dalam Bahasa Sasak | 16 |
| 2.3.4. Sistem Morfologi Bahasa Sasak | 19 |
| 2.3.5. Perbandingan Bahasa Indonesia dan Bahasa Sasak | 22 |
| 2.3.6. Pola Kalimat Bahasa Indonesia dan Bahasa Sasak | 25 |
| 2.3.7. Stemming | 28 |
| 2.3.8. Algoritma Nazief-Adriani | 29 |
| 2.3.9. Porter Stemmer | 32 |
| 2.3.10. Jaro-Winkler | 34 |
| BAB III METODE PENELITIAN | 38 |
| 3.1. Jenis, Sifat dan Pendekatan Penelitian | 38 |
| 3.1.1. Jenis Penelitian | 38 |
| 3.1.2. Sifat Penelitian | 38 |
| 3.1.3. Pendekatan Penelitian | 38 |
| 3.2. Metode Pengumpulan Data | 38 |
| 3.2.1. Studi literatur atau kepustakaan | 39 |
| 3.2.2. Observasi | 39 |
| 3.3. Metode Analisis Data | 39 |
| 3.4. Alur Penelitian | 39 |
| BAB IV HASIL PENELITIAN DAN PEMBAHASAN | 44 |
| 4.1. Pengumpulan Dan Normalisasi Data | 44 |
| 4.2. Modifikasi Algoritma | 45 |

| | |
|---------------------------|----|
| 4.3. Hasil Stemming | 47 |
| BAB V PENUTUP | 54 |
| DAFTAR PUSTAKA | 57 |
| Lampiran | 60 |

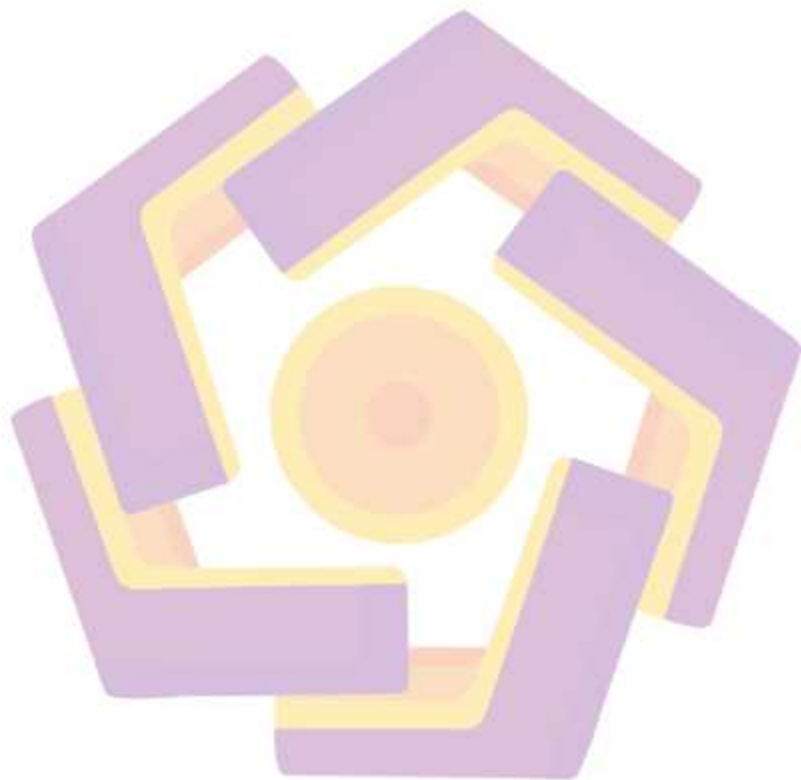


DAFTAR TABEL

| | |
|--|----|
| Tabel 2.1 Matriks literatur review dan posisi penelitian..... | 9 |
| Tabel 2.2 Perbandingan antar dialek..... | 18 |
| Tabel 2.3 Contoh Prefiks..... | 20 |
| Tabel 2.4 Contoh Infiks..... | 21 |
| Tabel 2.5 Contoh Sufiks..... | 21 |
| Tabel 2.6 Contoh Simufiks..... | 22 |
| Tabel 4.1 Hasil kumpulan kata dasar bahasa sasak..... | 44 |
| Tabel 4.2 Hasil Stemming Kombinasi Nazief-Adriani dan Porter Stemmer Beserta Jaro-Winkler Distance..... | 49 |
| Tabel 4.3 Hasil Stemming Kombinasi Nazief-Adriani dan Porter Stemmer Tanpa Jaro-Winkler Distance..... | 49 |
| Tabel 4.5 Hasil Stemming Kombinasi Nazief-Adriani dan Porter Stemmer Tanpa Jaro-Winkler Distance dan Tanpa Dictionary..... | 50 |
| Tabel 4.6 Hasil Stemming Nazief-Adriani Tanpa Jaro-Winkler Distance..... | 51 |
| Tabel 4.7 Perbandingan Tingkat keberhasilan <i>stemming</i> kata pada Bahasa Sasak | 52 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 3.1. Gambar Alur Penelitian | 41 |
| Gambar 4.1. <i>Pseudocode</i> penerapan similarity | 48 |



INTISARI

Perkembangan teknologi informasi saat ini ternyata mempunyai arti yang sangat penting dalam sentiment analysis, terutama kemampuan yang baik dalam mengolah data dan informasi mengenai opini masyarakat. Baik pemerintah ataupun perusahaan swasta tentu ingin mengetahui opini masyarakat agar bisa mengembangkan proyek atau rencana mereka kedepan. Dengan teknologi saat ini opini masyarakat sudah bisa diekstraksi menggunakan text mining di social media, dengan begitu pihak lain (pemerintah ataupun perusahaan swasta) dapat mengetahui opini masyarakat. Opini masyarakat Sasak dapat diekstraksi dengan algoritma yang sudah ada dapat dilakukan stemming tetapi ketika algoritma tersebut diaplikasikan pada bahasa yang berbeda bisa saja hasil yang diinginkan tidak sesuai. Penelitian ini bertujuan untuk melakukan stemming pada Bahasa Sasak menggunakan kombinasi algoritma Nazief-Adriani dan Porter Stemmer serta pengecekan similarity pada kata-kata yang luluh pada kamus data menggunakan algoritma Jaro-Winkler Distance.

Kata kunci: Stemmer, Porter Stemer, Nazief-Adriani, Jaro-Winkler Distance, Bahasa Sasak.



ABSTRACT

The development of information technology nowadays has a very important meaning in sentiment analysis, especially good ability to process data and information about public opinion. Both the government and private companies certainly want to know public opinion in order to develop their projects or plans going forward. With current technology, public opinion can be extracted using text mining in social media, so that other parties (government or private companies) can find out public opinion. Sasak community opinion can be extracted with existing algorithms stemming but when the algorithm is applied to a different language the desired results may not be appropriate. This research aims to stem Sasak language using a combination of Nazief-Adriani algoritma and Porter Stemmer and checking similarity of melted words in the data dictionary using the Jaro-Winkler Distance algorithm.

Keywords: Stemmer, Porter Stemer, Nazief-Adriani, Jaro-Winkler Distance, Sasak Language.



BAB I PENDAHULUAN

1.1. Latar Belakang Masalah

Indonesia merupakan negara yang kaya dengan budaya dan bahasa. Sampai dengan tahun 2017 sebanyak 668 bahasa dari 2.468 daerah di Indonesia telah teridentifikasi oleh Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan (Kemendikbud). Bahasa Sasak merupakan bahasa ibu yang dituturkan oleh suku Sasak yang menjadi etnis mayoritas di pulau Lombok, Indonesia. Bahasa ini berkerabat dekat dengan bahasa Bali dan bahasa Sumbawa yang dituturkan di pulau-pulau sekitar Lombok. Ketiganya merupakan bagian dari rumpun bahasa Austronesia. Bahasa Sasak tidak memiliki status resmi; di Indonesia bahasa Indonesia adalah bahasa nasional yang digunakan oleh penutur bahasa Sasak dalam konteks formal dan tertulis. Beberapa dialek bahasa Sasak memiliki tingkat kesalingpahaman yang rendah. Bahasa Sasak mempunyai sistem tingkatan bahasa, mirip dengan bahasa Jawa dan Bali. Setiap tingkatannya memiliki kosakata berbeda; penggunaannya ditentukan oleh status sosial relatif penutur terhadap lawan bicaranya. Meski kini jarang ditemui dalam ragam tulisan, teks-teks tradisional bahasa Sasak yang ditulis dengan medium lontar terkadang dibacakan pada acara-acara adat tertentu.

Pada umumnya, bagi masyarakat kebanyakan di Lombok, hanya dikenal dua bentuk bahasa dalam komunikasi sehari-hari, yaitu yang disebut dengan bahasa Sasak *biase/jamaq* atau *aok-ape* (ya-apa) dan Sasak *alus* atau *tiang-enggih* (saya-ya). Adapun bahasa Sasak sangat halus, yang disebut *kaji-meran* (saya-ya),

hanya dipakai oleh para datu-raden (raja dan kaum perwangsa atau ningrat). Klasifikasi itu didasarkan pada stratifikasi sosial masyarakat Sasak sebagai bangsawan atau menak (perwangsa) dan bukan bangsawan atau non-menak (Wiliam, 2006). Pada Bahasa Sasak juga terdapat beberapa dialek yaitu: dialek Pujut (a-e), dialek Bayan(a-a), dialek Selaparang (e-e) dan dialek Aiq Bukaq (a-o) (Hakim, 2016)

Perkembangan Bahasa Sasak terus berlanjut, layaknya Bahasa Indonesia yang akhirnya mempunyai Bahasa gaul. Jika membahas malah perkembangan, selalu ada yang paling cepat dalam berkembang, yaitu teknologi. Perkembangan teknologi informasi saat ternyata mempunyai arti yang sangat penting dalam *sentiment analysis*, terutama kemampuan yang baik dalam mengolah data dan informasi mengenai opini masyarakat. Baik pemerintah ataupun perusahaan swasta tentu ingin mengetahui opini masyarakat agar bisa mengembangkan proyek atau rencana mereka kedepan.

Internet atau sering disebut dengan dunia maya juga berkembang dengan pesat, dunia maya akhir-akhir ini digunakan untuk sarana jejaring sosial. Tidak sedikit masyarakat menuangkan opininya pada social media. Dengan teknologi saat ini opini masyarakat sudah bisa diekstraksi menggunakan text mining di social media, dengan begitu pihak lain (pemerintah ataupun perusahaan swasta) dapat mengetahui opini masyarakat.

Untuk mendapatkan opini masyarakat tersebut, perlu dilakukan stemming pada data-data yang ada di social media. Oleh karena itu peneliti mengembangkan algoritma untuk stemming bahasa Indonesia. Bahkan terdapat penelitian bertujuan

meningkatkan hasil stemming bahasa Indonesia untuk bahasa gaul (slang) (Maylawati, 2018).

Untuk bahasa Bali dalam penelitian yang menggunakan algoritma Porter Stemmer (Nata, 2017) dan penelitian yang menggunakan pendekatan rule-based dan metode N-gram stemming (Subali, 2019). Dengan adanya penelitian tersebut opini masyarakat dapat diekstrak dalam berbagai bahasa.

Opini masyarakat Sasak dapat diekstraksi dengan algoritma yang sudah ada dapat dilakukan stemming tetapi ketika algoritma tersebut diaplikasikan pada bahasa yang berbeda bisa saja hasil yang diinginkan tidak sesuai, oleh karena itu penelitian ini diajukan untuk stemming bahasa Sasak.

Penelitian ini dilakukan berdasarkan banyaknya penelitian terdahulu tentang data mining dan stemming, tetapi belum ada penelitian untuk bahasa Sasak. Selain agar lebih banyak mengenai penelitian terdahulu juga menjadikan bahan perbandingan metode manakah yang memiliki evaluasi terbaik.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan dalam penelitian ini dirumuskan masalah-masalah sebagai berikut:

- a. Bagaimana menerapkan stemming untuk bahasa Sasak?
- b. Berapa tingkat akurasi kombinasi algoritma Nazief-Adriani dan Porter Stemmer pada bahasa Sasak?
- c. Bagaimana pengaruh kombinasi antara algoritma Nazief-Adriani dan Porter Stemmer untuk stemming bahasa Sasak?

1.3. Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

- a. Tingkat tutur bahasa sasak yang distemming adalah bahasa sasak *jamaq* dan bahasa sasak *alus* tidak termasuk bahasa gaul atau kata-kata kombinasi yang sedang trend. Teks didapatkan dari teks cerita yang terdapat pada buku muatan local Gumi Sasak untuk SD/MI.
- b. Dialek bahasa Sasak yang distemming adalah dialek Pujut (a-e)
- c. Kombinasi stemmer yang digunakan adalah dari algoritma Nazief-Adriani dan Porter Stemmer
- d. Data kata dasar yang digunakan sebanyak 2.532 kata dari buku Tesaurus Bahasa Sasak (Nuryati, 2016)
- e. Algoritma tidak memproses istilah
- f. Hasil yang diukur dari algoritma adalah tingkat akurasi hasil stemming bahasa Sasak

1.4. Tujuan Penelitian

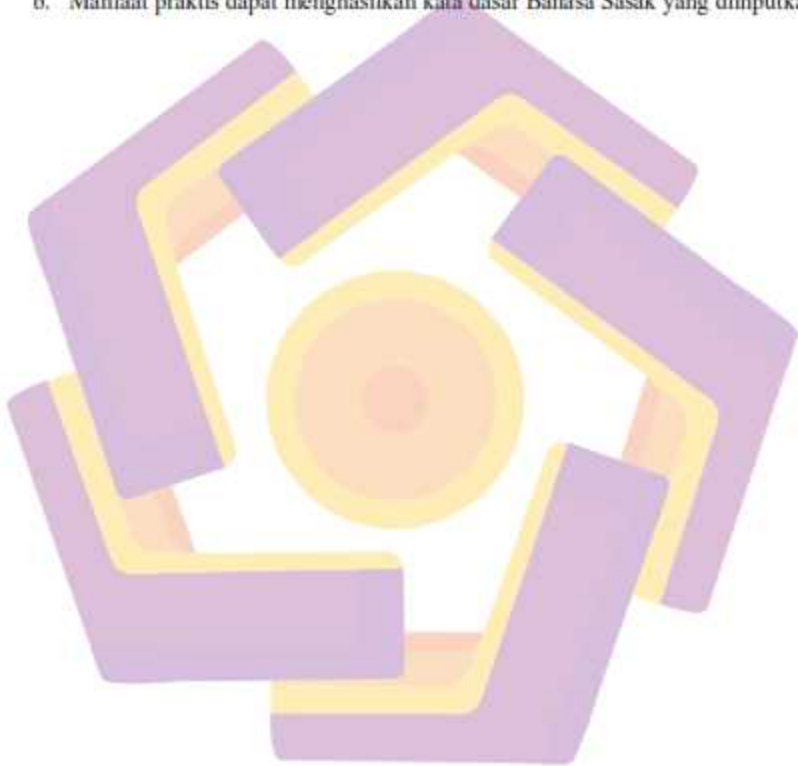
Tujuan dari penelitian ini adalah:

- a. Mendapatkan metode yang sesuai untuk stemming bahasa sasak
- b. Mengetahui tingkat akurasi kombinasi algoritma Nazief-Adriani dan Porter Stemmer dalam melakukan stemming untuk bahasa Sasak
- c. Mengetahui pengaruh kombinasi algoritma antara Nazief-Adriani dan Porter Stemmer dalam melakukan stemming untuk bahasa Sasak

1.5. Manfaat Penelitian

Manfaat yang dapat diambil dari penelitian ini adalah:

- a. Manfaat Teoritis menerapkan metode kombinasi algoritma untuk stemming bahasa Sasak
- b. Manfaat praktis dapat menghasilkan kata dasar Bahasa Sasak yang diinputkan



BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Stemming sendiri merupakan topik yang banyak diteliti. Sudah banyak penelitian mengangkat topik mengenai stemming. Stemming dapat dilakukan dengan algoritma yang sama maupun berbeda-beda untuk tiap bahasa. Pada penelitian terdahulu telah dikembangkan algoritma Nazief & Adriani ini yang menghasilkan akurasi dari 93% menjadi 50% (Asian, 2007).

Pada penelitian stemming Bahasa Bali menggunakan pendekatan metode Porter Stemmer untuk mengembangkan algoritma stemming khusus untuk Bahasa Bali dalam menangani sor-singgih pada dokumen bahasa Bali. Dalam penelitian tersebut menjelaskan adanya 2 langkah yang dilakukan peneliti. Pertama stemming, pemotongan kata untuk mendapatkan kata dasar bahasa Bali. Kemudian langkah selanjutnya adalah menerjemahkan ke Bahasa Indonesia menggunakan mapping kata dasar bali-indonesia yang sudah disimpan dalam database (Nata, 2017).

Penelitian stemming pada bahasa Bali bertujuan mengembangkan metode stemmer yang meluluhkan seluruh variasi afiks pada bahasa Bali dengan mengombinasikan pendekatan rule-based dan metode N-gram stemming. Untuk kesepuluh query metode yang diusulkan memperoleh rerata akurasi stemming lebih baik 96,67% dari metode terdahulu 75%, sedangkan untuk kelima query

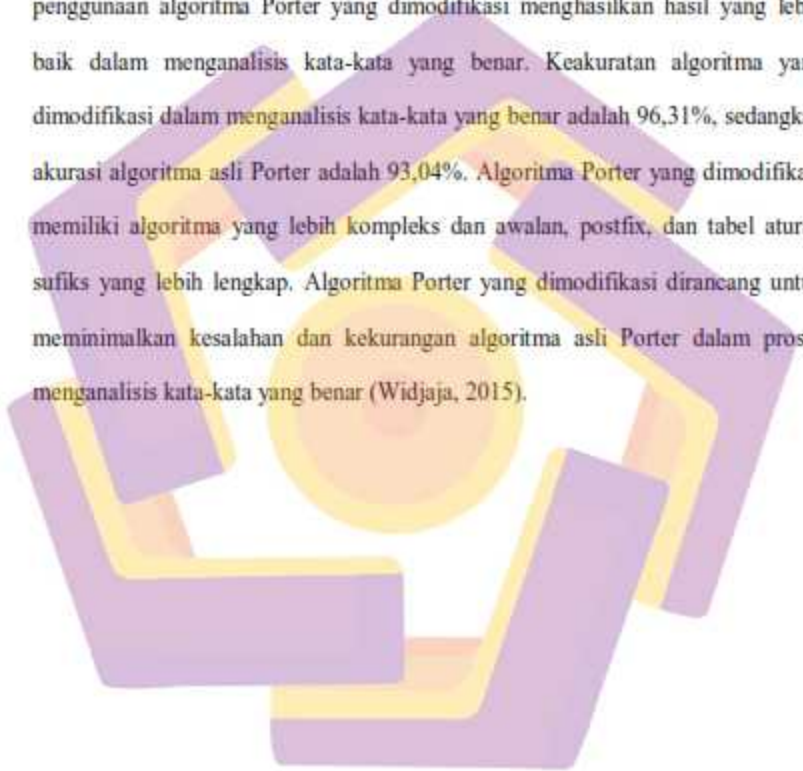
metode n-gram stemming dapat mengenali beberapa kata berafiks diluar rules (Subali, 2019).

Penelitian dengan pendekatan metode Nazief-Adriani kemudian memodifikasi Flexible Affix Classification (algoritma Nazief & Adriani) agar bisa melakukan stemming pada bahasa Indonesia non-formal. Hasil stemming dari 60 kata imbuhan non-formal yang dibandingkan dengan algoritma Klasifikasi Affiks Fleksibel dengan algoritma Affix Non-formal menunjukkan bahwa algoritma Affix Non-formal lebih akurat dalam membendung kata-kata imbuhan non-formal. Hasil tes menunjukkan bahwa algoritma Affix Nonformal masih tidak dapat melakukan stemming ke seluruh kata yang ditempelkan non-formal dalam data uji (Putra, 2018).

Kemudian Widayanto (2017) melakukan komparasi akurasi hasil stemming dengan metode Nazief-Adriani dan Confix Stemmer (CS). Hasil penelitian tersebut menunjukan bahwa adanya kemajuan besar dari algoritma Nazief-Adriani ke algoritma CS dengan menambahkan beberapa prefix dan modifikasi prefix yang sudah ada.

Dian (2018) berupaya meningkatkan hasil stemming Bahasa Indonesia untuk bahasa gaul (slang). Penelitian ini menggunakan algoritma stemming baru menggunakan Porter Stemmer sebagai acuan. Hasil stemming menggunakan algoritma baru tersebut dibandingkan dengan hasil algoritma Porter Stemmer original, Nazief-Adriani dan Luccene Stemmer dengan hasil yang terbukti meningkat menjadi 88,65%

Satu lagi penelitian yang memodifikasi Porter Stemmer dikembangkan khusus untuk plugin error detector. Algoritma asli Porter dan algoritma modifikasi Porter mampu menganalisis dengan sempurna semua kata yang salah (100%), namun, masih ada kekurangan dalam menganalisis kata-kata yang benar. penggunaan algoritma Porter yang dimodifikasi menghasilkan hasil yang lebih baik dalam menganalisis kata-kata yang benar. Keakuratan algoritma yang dimodifikasi dalam menganalisis kata-kata yang benar adalah 96,31%, sedangkan akurasi algoritma asli Porter adalah 93,04%. Algoritma Porter yang dimodifikasi memiliki algoritma yang lebih kompleks dan awalan, postfix, dan tabel aturan sufiks yang lebih lengkap. Algoritma Porter yang dimodifikasi dirancang untuk meminimalkan kesalahan dan kekurangan algoritma asli Porter dalam proses menganalisis kata-kata yang benar (Widjaja, 2015).



2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian

| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
|----|---------------------------------------|---|--|--|--|---|
| 1 | Stemming teks sor-singgih Bahasa Bali | Gusti Ngurah Mega Nata,Putu Pande Yudiastira, Konferensi Nasional Sistem & Informatika 2017. 2017 | Pengembangan algoritma stemming khusus untuk Bahasa Bali dalam menangani sor-singgih pada dokumen bahasa Bali. | Terdapat 2 langkah yang dilakukan peneliti pertama stemming, pemotongan kata untuk mendapatkan kata dasar bahasa bali. kemudian langkah selanjutnya adalah menerjemahkan ke bahasa indonesia menggunakan mapping kata dasar bali-indonesia yang sudah disimpan dalam database. | Mapping kata dasar bahasa bali tidak ada keterangan level bahasa. Sistem yang dibangun hanya mampu mencari kata dasar yang berisi awalan dan (atau) akhiran, sehingga belum bisa untuk kata yang berisi sisipan. | Pada penelitian yang akan diteliti stemming akan dilakukan untuk bahasa Sasak |

Tabel 2.1. Matriks literatur review dan posisi penelitian (Lanjutan)

| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
|----|---|---|--|--|---|--|
| 2 | Comparison Nazief Adriani And CS Stemmer Algorithm For Stem Real Data | Hari Widayanto, Arief Fatchul Huda. e-Proceeding of Engineering : Vol.4, No.3 Desember 2017, 2017 | Perbandingan akurasi hasil stemming dengan metode Nazief-Adriani dan Confix Stemmer (CS) | Adanya kemajuan besar dari algoritma Nazief-Adriani ke algoritma CS dengan menambahkan beberapa prefix dan modifikasi prefix yang sudah ada | Kurangnya keberagaman sample kata yang distemming, saat ini hanya dalam konteks berita. | Penelitian yang akan dilakukan stemming dalam menggunakan Porter Stemmer terlebih dahulu kemudian akan dilanjutkan dengan algoritma Nazief-Adriani |
| 3 | An Improved of Stemming Algorithm for Mining Indonesian Text with Slang on Social Media | Dian Sa'adillah Maylawati, Wildan Budiawan Zulfikar, Cepy Slamet, Muhammad Ali Ramdhani, Yana Aditia Gerhana. The 6th International Conference on Cyber and IT Service Management (CITSM 2018), 2018. | Meningkatkan hasil stemming bahasa indonesia untuk bahasa gaul (slang) | Hasil stemming menggunakan algoritma baru ini dibandingkan dengan hasil algoritma Porter Stemmer, Nazief-Adriani dan Luccene Stemmer dengan hasil yang terbukti meningkat menjadi 88,65% | Algoritma belum dapat menangani semua kasus awalan, terutama untuk awalan yang ditambahkan dengan huruf asimilasi. Dan penggunaan algoritma memori dapat dikurangi sehingga akan lebih efisien. | Penelitian yang akan dilakukan stemming dalam menggunakan Nazief-Adriani terlebih dahulu kemudian akan dilanjutkan dengan algoritma Porter Stemmer pada bahasa Sasak |
| 4 | Implementation of Porter's Modified Stemming Algorithm in an | Marsel Widjaja, Seng Hansun. International Journal of | mengembangkan, meningkatkan dan mengimplementasikan algoritma Porter temmer | Algoritma asli Porter dan algoritma modifikasi Porter mampu menganalisis | Algoritma Porter yang sudah dimodifikasi bisa digabungkan dengan algoritma lain supaya | Penelitian yang akan dilakukan stemming dalam menggunakan Nazief-Adriani terlebih dahulu kemudian |

Tabel 2.1. Matriks literatur review dan posisi penelitian (Lanjutan)

| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
|----|--|--------------------------------------|--|--|--|--|
| | Indonesian Word Error Detection Plugin Application | Technology (2015) 2: 139-150, 2015 | pada plugin pendeteksi kesalahan kata. | dengan sempurna semua kata yang salah (100%), namun, masih ada kekurangan dalam menganalisis kata-kata yang benar. penggunaan algoritma Porter yang dimodifikasi menghasilkan hasil yang lebih baik dalam menganalisis kata-kata yang benar. Keakuratan algoritma yang dimodifikasi dalam menganalisis kata-kata yang benar adalah 96,31%, sedangkan akurasi algoritma asli Porter adalah 93,04%. Algoritma Porter yang dimodifikasi memiliki algoritma yang lebih kompleks dan awalan, postfix, dan tabel aturan sufiks yang lebih lengkap. Algoritma Porter yang | menghasilkan output yang lebih akurat. | akan dilanjutkan dengan algoritma Porter Stemmer pada bahasa Sasak |

Tabel 2.1. Matriks literatur review dan posisi penelitian (Lanjutan)

| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
|----|--|--|--|--|--|--|
| | | | | dimodifikasi dirancang untuk meminimalkan kesalahan dan kekurangan algoritma asli Porter dalam proses menganalisis kata-kata yang benar. | | |
| 5 | Kombinasi Metode Rule-Based Dan N-Gram Stemming Untuk Mengenal Stemmer Bahasa Bali | Made Agus Putra Subali, Chastine Faticah. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) Vol. 6, No. 2, April 2019, 2019 | Pengembangan metode stemmer yang meluhukan seluruh variasi afiks pada bahasa Bali dengan mengombinasikan pendekatan rule-based dan metode n-gram stemming. | Untuk query metode yang diusulkan memperoleh rerata akurasi stemming lebih baik 96,67% dari metode terdahulu 75%, sedangkan untuk kelima query metode n-gram stemming dapat mengenali beberapa kata berafiks dituar rules. | Semantik setiap kata dan tahap validasi masih belum diperhatikan | Bahasa Bali dan bahasa Sasak dapat dikatakan <i>sister language</i> karena mempunyai rumpun yang sama, walaupun terdapat persamaan tentunya terdapat perbedaan dalam dua bahasa tersebut. Ada beberapa rule yang sama pada dua bahasa tersebut. Tetapi dalam penelitian ini tidak menggunakan rule-based dan hanya mengecek morfologi. |
| 6 | Non-formal Affixed Word Stemming in Indonesian Language | Rahardyan Bisma Setya Putra, Ema Utami. International Conference on Information and Communications | Modifikasi Flexible Affix Classification (algoritma Nazief & Adriani) agar bisa melakukan stemming pada bahasa Indonesia non-formal | Hasil stemming dari 60 kata imbuhan non-formal yang dibandingkan dengan algoritma Klasifikasi Afiks Fleksibel Affix dengan algoritma Affix | Tambahan pendekatan kesamaan untuk melakukan stemming pada kata imbuhan non-formal yang telah sedikit mengubah kata dasar. | Penelitian yang akan diteliti stemming akan dilakukan untuk bahasa Sasak |

Tabel 2.1. Matriks literatur review dan posisi penelitian (Lanjutan)

| No | Judul | Peneliti, Media Publikasi, dan Tahun | Tujuan Penelitian | Kesimpulan | Saran atau Kelemahan | Perbandingan |
|----|-------|--------------------------------------|-------------------|--|----------------------|--------------|
| | | Technology (ICOIACT). 2018 | | Non-formal menunjukkan bahwa algoritma Affix Non-formal lebih akurat dalam membendung kata-kata imbuhan non-formal. Hasil tes menunjukkan bahwa algoritma Affix Nonformal masih tidak dapat melakukan stemming ke seluruh kata yang ditempelkan non-formal dalam data uji. | | |

2.3. Landasan Teori

2.3.1. Bahasa Sasak

Sasak dan Lombok mempunyai hubungan yang sangat erat sehingga dua kata tersebut tidak bisa dipisahkan. Keduanya terjalin menjadi satu yang berasal dari kata "*saqsaq lomboq*". Kata *saq* berarti satu dan *lomboq* berarti lurus (Sudirman, 2012). Dengan demikian, Sasak Lombok berarti satunya lurus atau satu-satunya kelurusan. Dari segi bahasa, bahasa Sasak sangat sederhana, paling banyak terdiri atas dua suku kata. Hal ini terlihat pada kata-kata yang menunjukkan tempat yang cukup ditambahkan kata *timuq* "timur", *bat* "barat", *daye* "utara", dan *lauq* "selatan" (Sudirman, 2012). Orang-orang Sasak atau Lombok pun adalah orang-orang yang sederhana dan mahal berpakaian, bertutur kata, meniti kehidupan (Hakim, 2016).

Bahasa Sasak merupakan merupakan salah satu bahasa daerah yang digunakan dan dilestarikan oleh masyarakat pemakainya yaitu suku Sasak, suku asli masyarakat yang bertempat tinggal di Pulau Lombok dan pulau-pulau kecil yang berada di sekitarnya. Di samping itu, pemakaian Bahasa Sasak ada juga di masyarakat yang bertempat tinggal di daerah-daerah kantong transmigrasi, seperti di Pulau Sumbawa, Pulau Sulawesi, Pulau Kalimantan dan daerah-daerah lainnya. Jumlah penutur Bahasa Sasak di Provinsi Nusa Tenggara Barat sebanyak 2.594.629 orang (Sensus Penduduk Tahun 2000) dengan rincian 2.508.484 orang di pulau Lombok dan 86.145 orang di Pulau Sumbawa. Apabila jumlah tersebut dikaitkan dengan pandangan Ferguson (dalam Abbas, 1983), maka hal itu berarti

bahwa Bahasa Sasak termasuk bahasa mayor atau *major language* karena berpenutur lebih dari 1 juta orang (Hakim, 2016).

2.3.2. Tingkat Tutur Bahasa Sasak

Bahasa Sasak adalah bahasa suku Sasak. Pada umumnya, bagi masyarakat kebanyakan di Lombok, hanya dikenal dua bentuk bahasa dalam komunikasi sehari-hari, yaitu yang disebut dengan bahasa Sasak *biase/jamaq* atau *aok-ape* (ya-apa) dan Sasak *alus* atau *tiang-enggih* (saya-ya). Adapun bahasa Sasak sangat halus, yang disebut *kaji-meran* (saya-ya), hanya dipakai oleh para datu-raden (raja dan kaum perwangsa atau ningrat). Klasifikasi itu didasarkan pada stratifikasi sosial masyarakat Sasak sebagai bangsawan atau menak (perwangsa) dan bukan bangsawan atau non-menak (Wiliam, 2006).

Faktor yang menyebabkan pemilihan terhadap bentuk bahasa yang digunakan, seperti juga dalam bahasa lain, adalah usia, status sosial, pendidikan, tingkat keakraban, situasi percakapan, jenis percakapan (*formal/ informal*). Ada satu lagi tingkat tutur bahasa Sasak yang jarang digunakan, yaitu bahasa Sasak kasar. Tingkat tutur kasar merupakan *tingkatan* kebahasaan dalam bahasa Sasak yang digunakan ketika seseorang marah atau merasa tidak senang terhadap lawan bicara atau topik pembicaraan. Jumlah kosakata tingkat tutur kasar sangat sedikit.

Tutur kedua, bahasa *jamaq* atau bahasa pergaulan adalah bahasa yang paling lumrah atau bahasa sehari-hari yang digunakan oleh penutur bahasa Sasak pada umumnya. Tingkatan ini biasa digunakan oleh penutur kebanyakan, dan penutur yang memiliki keakraban. Tingkat tutur yang ketiga, bahasa *alus* (*madya*) biasanya digunakan untuk menghormati lawan bicara atau topik pembicaraan.

Tingkatan ini biasanya digunakan oleh penutur yang lebih muda atau lebih rendah status sosialnya kepada penutur yang lebih tua atau lebih tinggi status sosialnya. Hampir sama dengan tingkatan tutur yang ketiga, tingkat tutur yang keempat juga digunakan untuk menghormati lawan bicara atau topik pembicaraan. Tingkatan ini biasanya digunakan kepada orang yang sangat dihormati atau disegani. Perlu dicatat disini bahwa terjadi perubahan sosial dan linguistik dalam komunitas Sasak (Mahyuni, 2006), yaitu perubahan gaya bahasa yang dikenal dengan alus. Sekarang, alus bukan lagi identik/milik orang menak tapi bergeser ke penutur dengan gelar keagamaan, pendidikan, dan status sosial. (Hakim, 2016)

2.3.3. Dialek Dalam Bahasa Sasak

Rekonstruksi Bahasa Sasak dijelaskan Mahsun (2006) sebagai berikut. Secara kualitatif, hubungan kekerabatan di antara keempat dialek Bahasa Sasak dinyatakan dengan adanya hubungan dialek yang meneruskan satu bahasa induk, Prabahasa Sasak. Secara historis Dialek Pujut (DP) dan Dialek Selaparang(DS) diduga sebagai subdialek dari satu dialek, yaitu dialek Pujut-Selaparang (DPS). Dalam perkembangan Bahasa Sasak modern, kedua subdialek tersebut muncul sebagai dialek yang berdiri sendiri. Dialek Dialek Bayan (DB) merupakan dialek Bahasa Sasak yang lebih konservatif dan lebih tua sejarahnya dibandingkan dengan dialek yang lain. Selanjutnya, Dialek Aiq Bukaq (DA) merupakan variasi dialectal yang muncul akibat dari adanya inovasi eksternal yang dialami Bahasa Sasak, khususnya pengaruh dari Jawa (Hakim, 2016).

Dalam bahasa Sasak juga terdapat dialek yaitu (Hakim, 2016):

- a) Dialek Bayan (a-a), penutur dialek Bayan sebagian besar menyebar di daerah pegunungan Lombok Utara dan sedikit wilayah Lombok Barat, Tengan dan Timur. Secara fonologis dialek a-a mewarisi bentuk asli Austronesia.
- b) Dialek Pujut (a-e), penutur dialek ini menyebar dari bagian barat ke bagian tengah dan sedikit di bagian timur pulau Lombok. Dialek a-e terpengaruh bahasa Bali, hal ini terbukti dengan gaya bicara penuturnya yang kebalibalian.
- c) Dialek Aiq Bukaq (a-o), penuturnya menyebar di wilayah Lombok Tengah dan sebagian kecil di Lombok Timur. Dialek a-o terpengaruh bahasa Jawa. Hal ini disebabkan bahasa Jawa membawa pengaruh yang besar selama kerajaan Singosari dan Majapahit
- d) Dialek Selaparang (e-e), penuturnya tersebar di bagian tengah dan timur, dan sedikit di bagian barat pulau Lombok. Dialek e-e adalah dialek bahasa yang lahir sebagai symbol perlawanan secara lingual terhadap bahasa Bali karena pada saat itu Bali menjadi penguasa di pulau Lombok.

Perbandingan antar dialek pada Tabel 2.2. Dari tabel tersebut dapat kita amati bahwa kosakata dari berbagai dialek yang ada terdapat perbedaan yang sangat kecil dan dapat dipandang sebagai perbedaan dalam pengucapan sehari-hari.

Tabel 2.2. Perbandingan antar dialek

| NO | Kosakata | | | | Arti |
|----|----------|------|------|------|--------|
| | a-e | e-e | a-a | a-o | |
| 1 | Dade | Dede | Dada | Dado | Dada |
| 2 | Mate | Mete | Mata | Mato | Mata |
| 3 | Ape | Epe | Apa | Apo | Apa |
| 4 | Pire | Pire | Pira | piro | Berapa |

Di antara empat dialek Bahasa Sasak (dialek a-e atau dialek Pujut, dialek e-e atau dialek Selaparang, dialek a-a atau dialek Bayan, dan dialek a-o atau dialek Aiq Bukaq) yang paling memungkinkan untuk dijadikan bahasa Sasak standar adalah dialek a-e (Dialek Pujut) (Sirulhaq, 2009). Hal ini berdasarkan pertimbangan-pertimbangan sebagai berikut (Hakim, 2016):

- a) Secara Geografis. Sebaran penutur dialek a-e (Dialek Pujut) dibandingkan dengan dialek- dialek yang lain lebih luas, pada daerah-daerah yang memiliki kepadatan penduduk yang relatif tinggi. Di samping itu sebarannya mencakup wilayah kota Mataram yang relatif lebih mudah dan cenderung ditiru karena dianggap lebih memiliki nilai prestisius
- b) Secara Linguistik. Dialek a-a(Dialek Bayan) dan dialek e-e(Dialek Selaparang) cenderung memiliki logat yang sama dan demikian pula sama dalam pola urutan kata (word order) dalam kalimat.

Sedangkan dialek a-e(Dialek Pujut) cenderung berbeda dengan dua dialek di atas dari segi logat dan pola urutan kata.

- c) Secara Sociolinguistic. Para penutur Bahasa Sasak menilai bahwa dialek a-e(Dialek Pujut) merupakan dialek yang paling banyak diikuti dan dianggap berlaku umum. Berkelakuan umum yang dimaksud disini bahwa rata-rata penutur masing-masing dialek memahami bahasa dialek a-e(Dialek Pujut), sebaliknya tidak semua penutur Bahasa Sasak memahami dengan baik dialek selain dialek tersebut. Adanya pengaruh bahasa halus dalam dialek a-e(Dialek Pujut) membawa citra tersendiri terhadap dialek ini. Citra inilah yang menyebabkan dialek lain selain dialek a-e(Dialek Pujut) dianggap sebagai dialek yang kasar, tidak patut diajarkan di sekolah. Program-program acara lokal di media massa, seperti radio, koran, dan televisi diisi dengan dialek a-e(Dialek Pujut) karena dialek ini dianggap familier dan positif. Begitu juga dengan lagu-lagu Sasak, hampir semuanya menggunakan struktur dialek a-e (Dialek Pujut).

2.3.4. Sistem Morfologi Bahasa Sasak

Morfologi adalah telaah secara structural terhadap morfem-morfem beserta penyusunannya dalam rangka pembentukan kata yang banyak terpakai dalam bahasa Sasak. Morfem dalam bahasa Sasak terbagi menjadi dua, yaitu morfem bebas dan morfem terikat. (Hakim, 2016)

Morfem bebas dalam penggunaan bahasa Sasak ada beberapa contoh, contoh-contoh ini dapat dikatakan morfem bebas karena dapat berdiri sendiri dan dapat diucapkan tersendiri walaupun tidak diletakkan dalam hubungan kalimat:

- a. Satu suku kata : saq(yang atau satu), to(disana), leq(di – keterangan tempat), jak(akan), yaq(akan), gen(akan)
- b. Dua suku kata : bale(rumah), bareng(bersama), awis
- c. Tiga suku kata : beriuk(bersama-sama)

Morfem terikat adalah morfem yang tidak dapat berdiri sendiri, mengandung makna setelah dipadukan dengan morfem lain atau bentuk lain. Morfem terikat terbagi menjadi dua yaitu morfem yang terikat secara morfologis dan morfem terikat secara sintaksis. Dalam penelitian ini akan dibahas tentang morfem terikat secara morfologis. Morfem terikat secara morfologis terbagi menjadi 3(Hakim, 2016), yaitu;

- a. Prefiks (awalan)

Awalan adalah morfem yang terletak di depan kata yang mengikatnya, misalnya /be-/ dalam kata *begawe*. Berikut terdapat contoh pada Table 2.3.

Table 2.3. Contoh Prefiks

| Prefiks | Morfem terikat |
|---------|--|
| /be-/ | Begawe, bekedek, bedait, beruni, berujuk |
| /ber-/ | Berobah, berongkos, beradat |
| /pe-/ | Pemaling, penyopet, penguinem |
| /peng-/ | Pengkedek, pengawis |
| /me-/ | Menyusah, memaling, meliwat |

| Prefiks | Morfem terikat |
|---------|------------------------------|
| /nge-/ | Ngeraos, ngengakoq, ngengais |

b. Infiks (sisipan)

Beberapa infiks produktif dalam bahasa Sasak terdapat pada Table 2.4.

Table 2.4. Contoh Infiks

| Infiks | Morfem Terikat |
|--------|---------------------|
| /-er-/ | Geruduh, geramus |
| /-el-/ | Gelompong, belunjur |
| /-eg-/ | Gegitaq, gegoloq |
| /-em-/ | Pemaling, pemujiq |

c. Sufiks (akhiran)

Beberapa akhiran produktif dalam bahasa Sasak terdapat pada Table 2.5.

Table 2.5. Contoh Sufiks

| Sufiks | Morfem Terikat |
|--------|--------------------------------|
| /-an/ | Piyaan, baitan |
| /-ang/ | Kadu-ang, gulah-ang, kodeq-ang |
| /-n/ | Beleq-n, yaq-n, jaq-n |
| /-m/ | Neq-m, yaq-m, mele-m |
| /-ne/ | Anuq-ne, piyaq-ne, serio-ne |
| /-ku/ | Anuq-ku, yaq-ku |

d. Simufliks (Kombinasi awalan dan akhiran)

Banyak morfem terikat yang terbentuk dari kombinasi antara prefix dan sufiks atau kombinasi ketiganya. Beberapa simufliks produktif dalam bahasa Sasak terdapat pada Table 2.6.

Table 2.6. Contoh Simufliks

| Sufiks | Morfem Terikat |
|---------|---------------------------|
| /be-an/ | Be-gawe-an, be-jagur-an |
| /pe-an/ | Pe-rage-an, pe-kedek-an |
| /ke-an/ | Ke-jari-an, ke-kedek-an |
| /te-an/ | Te-talet-an, te-tangis-an |

2.3.5. Perbandingan Bahasa Indonesia dan Bahasa Sasak

Terdapat beberapa aspek untuk mengidentifikasi persamaan dan perbedaan antara dua bahasa. Dalam hal ini adalah perbandingan antara Bahasa Indonesia dan Bahasa Sasak yang dijabarkan dalam point-point berikut dalam sebuah skripsi analisis kontrastif sintaksis Bahasa Indonesia dan bahasa sasak dialek a-e (Baiq, 2013);

a. Perpadanan (*equivalence*)

Perpadanan adalah hubungan yang terjadi antara kalimat dalam suatu bahasa dengan kalimat dalam bahasa lain hanya apabila masing-masing kalimat itu merupakan hasil penerjemahan optimal dari yang lainnya dalam konteks tertentu. Jadi, padanan suatu kata atau frasa dalam kalimat adalah kata

atau frasa itu pula yang berubah dalam kalimat bahasa lain dan yang mempunyai fungsi sintaksis yang sama.

Contoh : bahasa Sasak → Meri tendang lawang.

S P O

bahasa Indonesia → Meri menendang pintu.

S P O

Kalimat dalam bahasa Indonesia di atas merupakan penerjemahan optimal dari kalimat bahasa Sasak, begitu pula sebaliknya. Pada kalimat tersebut, masing-masing fungsi subjek diduduki oleh nomina ; verba *tendang* 'menendang' sama-sama menduduki fungsi predikat, dan *lawang* 'pintu' menduduki fungsi yang sama, yakni objek. Jadi, kalimat bahasa sasak "*Meri tendang lawang*" adalah perpadanan dari kalimat bahasa Indonesia 'Meri menendang pintu'.

b. Kesejajaran Bentuk (*formal correspondance*)

Kriteria penilaian yang kedua ialah kesejajaran bentuk. Kesejajaran bentuk merupakan persamaan struktur antara dua bahasa yang terdapat dalam berbagai tingkat atau tataran.

Contoh : bahasa Sasak → Saiq lalo jok bangket.

bahasa Indonesia → Bibi pergi ke sawah.

S P Ket. tempat

Subjek *saiq* 'bibi' pada pasangan kalimat di atas sama-sama berbentuk nomina, predikat *lalo* 'pergi' berbentuk verba, dan keterangan tempat *jok bangket* 'ke sawah' merupakan frasa preposisional. Jadi, kalimat bahasa Sasak

“*Saiq lalo jok bangket*” dan kalimat bahasa Indonesia ‘Bibi pergi ke sawah’ memiliki kesejajaran bentuk.

c. Kesamaan Urutan Kata (*identical word order*)

Menurut Syamsun (1992:23), istilah urutan kata yang digunakan oleh Marton memiliki pengertian yang umum. Oleh sebab itu, dia mengartikannya sebagai urutan unsur pengisi fungsi sintaksis.

Contoh : bahasa Sasak → Inaq ronas piring leq sumur.

 bahasa Indonesia → Ibu mencuci piring di sumur.

Unsur *Inaq* ‘Ibu’ pada kedua kalimat di atas, yang berfungsi sebagai subjek, terletak sebelum predikat *ronas* ‘mencuci’, kata *piring* ‘piring’ sebagai objek menempati urutan setelah predikat, dan terakhir frasa preposisional *leq sumur* ‘di sumur’ yang berfungsi sebagai keterangan tempat sama-sama berada di akhir kalimat.

d. Kesamabangunan (*congruence*)

Kesamabangunan adalah paduan dari perpadanan, kesejajaran bentuk, dan kesamaan urutan kata yang terjadi secara simultan antara kalimat-kalimat atau bagian-bagian kalimat dalam dua bahasa yang berbeda.

Contoh : Aku sapu lelah.

A B C

Saya menyapu halaman.

A' B' C'

Kelas kata :

N V N

Fungsi sintaksis:

S P O

Kalimat bahasa Sasak di atas (A, B, C) sebangun (~) dengan padanannya dalam bahasa Indonesia (A', B', C'). Hal ini karena masing-masing pasangan (AA', BB', CC') telah memenuhi syarat perpadanan, kesejajaran bentuk, dan kesamaan urutan kata. Selain itu, unsur-unsur yang berpadanan tersebut memiliki kesamaan kelas kata dan fungsi sintaksis di dalam kalimat.

2.3.6. Pola Kalimat Bahasa Indonesia dan Bahasa Sasak

Dalam KBBI (1993 : 434), kalimat diartikan sebagai kesatuan ujar yang mengungkapkan suatu konsep pikiran dan perasaan. Menurut Martinet (1987) dalam Syamsun (1992 : 32) kalimat adalah gabungan dua monem atau lebih yang predikatif dan sudah mempunyai makna yang lengkap. Sedangkan Chaer (2009:44) mendefinisikan kalimat sebagai satuan sintaksis yang disusun dari konstituen dasar yang biasanya berupa klausa, dilengkapi dengan konjungsi bila diperlukan, serta disertai dengan intonasi final. Lebih rinci lagi, di dalam buku Tata Bahasa Baku Bahasa Indonesia (Alwi, 2003 : 311) pengertian kalimat adalah satuan bahasa terkecil, dalam wujud lisan maupun tulisan yang mengungkapkan pikiran yang utuh. Dalam wujud lisan, kalimat diungkapkan dengan suara naik turun dan keras lembut, disela jeda, dan diakhiri dengan intonasi tanpa proses fonologis lainnya. Sedangkan dalam wujud tulisan, kalimat diawali dengan huruf kapital dan diakhiri dengan tanda titik, tanda tanya, atau tanda seru ; sementara itu, di dalamnya disertakan dengan tanda baca seperti koma, titik dua, tanda pisah, dan tanda spasi.

Tiap kata atau frasa dalam kalimat mempunyai fungsi yang mengaitkannya dengan kata atau frasa lain yang ada dalam kalimat tersebut. Fungsi itu bersifat

sintaksis, artinya berkaitan dengan urutan kata atau frasa dalam kalimat. Fungsi sintaksis utama adalah *predikat* (P), *subjek* (S), *objek* (O), *pelengkap* (Pel.), dan *keterangan* (Ket.) (Alwi, 2003:36). Urutan fungsi-fungsi tersebut ada yang harus tetap tetapi ada pula yang tidak tetap.

Contoh :

1. Meri berlari.
S P
2. Nenek membeli bubur.
S P O
3. Ia merengek kesakitan.
S P Pel.
4. a. Ia mengirim istrinya uang setiap bulan.
S P O Pel. Ket.
b. Setiap bulan ia mengirim istrinya uang.
Ket. S P O Pel.
5. a. Ayah memotong kue dengan perlahan.
S P O Ket.
b. Dengan perlahan ayah memotong kue.
Ket. S P O
6. Kemarin nenek meninggal.
Ket. S P
7. Di halaman ibu menanam bunga.
Ket. S P O

Kalimat bahasa Indonesia di atas masing-masing memiliki fungsi sintaksis yang menjadi unsur pembangun sebuah kalimat. Jika diperhatikan polanya, subjek selalu mendahului predikat, dan predikat selalu mendahului objek. Pola semacam itu (predikat terletak setelah subjek) merupakan pola umum kalimat dasar dalam

bahasa Indonesia (Alwi, 2003:322). Sementara itu, keterangan bisa pada awal atau akhir kalimat atau klausa (Chaer, 2009:33).

Berbeda halnya dengan pola kalimat bahasa Sasak yang letak predikatnya lebih fleksibel. Artinya, kalimat yang subjeknya mendahului predikat ataupun predikatnya mendahului subjek sama-sama berterima dalam bahasa ini.

Contoh :

1. a. Nyengke-'n mopoq (inaq) leq sumur. 'Ibu sedang mencuci (baju) di sumur'.
sedang - 3 P S Ket.
- b. Inaq nyengke mopoq leq sumur. 'Ibu sedang mencuci (baju) di sumur'.
S P Ket.
2. a. Sakit-'n (inaq). 'Ibu sakit'.
P - 3 S
- b. Inaq sakit. 'Ibu sakit'.
S P
3. a. Bedagang-'n (saiq) leq peken. 'Bibi berjualan di pasar'.
P - 3 S Ket.
- b. Saiq bedagang leq peken. 'Bibi berjualan di pasar'.
S P Ket.

Berdasarkan contoh di atas, terdapat perbedaan yang mendasar antara pola kalimat bahasa Sasak dengan bahasa Indonesia. Fungsi predikat dalam bahasa Indonesia tidak lazim ditemukan di awal kalimat, sedangkan di dalam bahasa Sasak distribusi fungsi tersebut bisa lebih fleksibel, bisa terletak sebelum atau setelah subjek.

Ada beberapa pengklasifikasian kalimat, di antaranya adalah berdasarkan jumlah klausa, fungsi, dan ada tidaknya unsur sangkalan (negasi) di dalam kalimat

tersebut. Berdasarkan jumlah klausanya, kalimat dibedakan menjadi kalimat tunggal dan kalimat majemuk. Berdasarkan fungsinya, ada jenis kalimat berita (deklaratif), perintah (imperatif), tanya (interogatif), dan seruan (interjektif). Sedangkan dari segi ada tidaknya unsur negasi, kalimat dibedakan menjadi kalimat positif dan kalimat negatif. Namun, yang akan dibahas dalam penelitian ini hanyalah kalimat tunggal deklaratif, yakni kalimat berita yang hanya mengandung satu klausa. Fokus tersebut disesuaikan dengan judul penelitian ini. Itulah sebabnya, kalimat-kalimat yang dimaksud bisa dijadikan sebagai data yang akan dianalisis.

2.3.7. Stemming

Proses untuk mengekstraksi kata dasar dari kata berafiks dikenal dengan istilah *stemming* (Balasankar, 2016) yang bertujuan meningkatkan recall dengan mereduksi variasi kata berafiks ke dalam bentuk kata dasarnya (Patil, 2017) (Pramudita, 2018).

Dalam penjelasan lain juga disebutkan bahwa *stemming* merupakan suatu metode bagian dari NLP (Putra, 2018) Stemming adalah suatu proses pelepasan afiks dari sebuah kata (Bird, 2014). Pada stemming biasanya terdapat error *understemming* dan *overstemming*. *Understemming* adalah ketika kata teral terlalu sedikit dipotong imbuhan, sedangkan *overstemming* adalah ketika kata terlalu banyak dipotong imbuhan (Kaara, 2013).

Stemming bukanlah proses yang selalu menghasilkan informasi yang optimal untuk setiap kasus, biasanya teknik stemming dipilih berdasarkan kebutuhan dan tujuan dari proses NLP itu sendiri (Bird, 2014). Proses

stemming biasanya dilakukan dengan cara menghilangkan sufiks dan prefix (afiks) dari suatu kata untuk mendapatkan kata dasarnya (Tari, 2016). Stemming dapat digunakan pada berbagai aplikasi pemrosesan teks seperti information retrieval, pengecekan plagiarisme, meningkatkan performa pencarian, dan lainnya (Putra, 2018)

Teknik ini memiliki algoritma yang berbeda-beda pada setiap bahasa. Pada Bahasa Inggris, algoritma porter menjadi algoritma yang banyak digunakan dan sudah seperti algoritma yang standar dalam melakukan stemming (Willet, 2006). Bahasa Indonesia memiliki beberapa algoritma stemming seperti Nazief & Adriani, Porter untuk Indonesia, Vega, dan Arifin & Setiono. Algoritma Nazief & Andriani memiliki akurasi paling baik berdasar penelitian yang sudah dilakukan sebelumnya (Asian, 2007). Algoritma stemming Nazief & Andriani banyak dikembangkan dan dimodifikasi. Algoritma ini menjadi basis algoritma stemming pada Bahasa Indonesia.

2.3.8. Algoritma Nazief-Adriani

Algoritma Nazief-Adriani adalah salah satu algoritma untuk stemming. Skema stemming dari Nazief dan Adriani dijelaskan dalam laporan teknis yang tidak dipublikasikan dari Universitas Indonesia pada tahun 1996. Algoritma ini didasarkan pada aturan morfologis komprehensif yang mengelompokkan dan merangkul imbuhan diizinkan dan tidak diizinkan, termasuk awalan, akhiran dan imbuhan (kombinasi awalan dan akhiran). Imbuhan dapat bersifat reflektif atau turunan (Asian, 2007)

Algoritma Nazief & Adriani dikembangkan pertama kali oleh Bobby Nazief dan Mirna Adriani. Algoritma ini berdasarkan pada aturan morfologi Bahasa Indonesia yang luas, yang dikumpulkan menjadi satu grup dan dienkapsulasi pada imbuhan/affixes yang diperbolehkan (allowed affixes) dan imbuhan/affixes yang tidak diperbolehkan (disallowed affixes). Algoritma ini menggunakan kamus kata dasar dan mendukung recoding, yakni penyusunan kembali kata-kata yang mengalami proses stemming berlebih. (Wahyudi, 2017)

Langkah-langkah algoritma Nazief & Adriani adalah:

- 1) Kata yang belum di-stemming dicari pada kamus, jika ditemukan, kata tersebut dianggap sebagai kata dasar yang benar dan algoritma dihentikan.
- 2) Hilangkan Inflectional suffixes, yaitu dengan menghilangkan particle (“-lah”, “-kah”, “-tah” atau “-pun”), kemudian hilangkan inflectional possessive pronoun suffixes (“-ku”, “-mu” atau “-nya”). Cek kata di dalam kamus kata dasar, jika ditemukan, algoritma dihentikan, jika tidak lanjut ke langkah 3.
- 3) Hapus Derivational Suffix (“-i” atau “-an”). Jika kata ditemukan dalam kamus kata dasar, maka algoritma berhenti. Jika tidak, maka lanjut ke langkah 3a:
 - a) Jika akhiran “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga dihapus. Jika

kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.

b) Akhiran yang dihapus (“-i”, “- an” atau “-kan”) dikembalikan, lanjut ke langkah 4.

4) Hapus Derivational Prefix (“be-”, “di-”, “ke-”, “me-”, “pe-”, “se-” dan “te-”). Jika kata yang didapat ditemukan didalam database kata dasar, maka proses dihentikan, jika tidak, maka lakukan recoding. Tahapan ini dihentikan jika memenuhi beberapa kondisi berikut:

- a) Terdapat kombinasi awalan dan akhiran yang tidak diijinkan
- b) Awalan yang dideteksi sama dengan awalan yang dihilangkan sebelumnya.
- c) Tiga awalan telah dihilangkan

5) Jika semua langkah telah dilakukan tetapi kata dasar tersebut tidak ditemukan pada kamus, maka algoritma ini mengembalikan kata yang asli sebelum dilakukan stemming.

Dalam penelitiannya Asian, dkk (Asian, 2005) melakukan beberapa pengembangan algoritma Nazief & Adriani sebagai berikut:

- 1) Menggunakan kamus kata yang lebih lengkap
- 2) Menambahkan aturan-aturan untuk kata-kata majemuk perulangan.

- 3) Menambahkan aturan awalan dan akhiran, serta aturan lainnya, yaitu:
 - a) Menambahkan partikel (inflection suffix) “-pun”.
 - b) Penambahan aturan pemenggalan awalan.
 - c) Perubahan aturan pemenggalan untuk tipe awalan “me”.
- 4) Perubahan urutan proses stemming, yaitu:
 - a) Kata dengan awalan “be-” dan akhiran “-lah”, hilangkan awalan terlebih dahulu kemudian akhiran.
 - b) Kata dengan awalan “be-” dan akhiran “-an”, hilangkan awalan terlebih dahulu kemudian akhiran.
 - c) Kata dengan awalan “me-” dan akhiran “-i”, hilangkan awalan terlebih dahulu kemudian akhiran.
 - d) Kata dengan awalan “di-” dan akhiran “-i”, hilangkan awalan terlebih dahulu kemudian akhiran.
 - e) Kata dengan awalan “pe-” dan akhiran “-i”, hilangkan awalan terlebih dahulu kemudian akhiran.
 - f) Kata dengan awalan “ter-” dan akhiran “-i”, hilangkan awalan terlebih dahulu kemudian akhiran.

2.3.9. Porter Stemmer

Algoritma Porter stemming adalah stemmer konflasi yang diusulkan oleh Porter. Algoritma ini didasarkan pada gagasan bahwa imbuhan dalam bahasa Inggris sebagian besar terdiri dari kombinasi imbuhan yang lebih kecil dan lebih sederhana. Proses pengupasan dilakukan pada serangkaian langkah, khususnya

lima langkah, yang mensimulasikan proses infektif dan turunan kata. Pada setiap langkah, imbuhan tertentu dihapus dengan aturan substitusi. Aturan substitusi diterapkan ketika seperangkat kondisi / kendala yang melekat pada aturan ini berlaku. Salah satu contoh kondisi seperti itu adalah panjang minimal (jumlah urutan konsonan-vokal) dari batang yang dihasilkan. Panjang minimum ini disebut ukuran. Kondisi sederhana lainnya pada batang dapat berupa apakah batang berakhir dengan konsonan, atau apakah batang mengandung vokal.

Ketika semua kondisi aturan tertentu terpenuhi, aturan diterapkan, yang menyebabkan penghapusan imbuhan dan kontrol bergerak ke langkah berikutnya. Jika kondisi aturan tertentu dalam langkah saat ini tidak dapat dipenuhi, kondisi aturan berikutnya dalam langkah tersebut diuji, sampai aturan tersebut selesai atau aturan dalam langkah itu habis. Proses ini berlanjut untuk semua lima langkah (Tala, 2002).

Algoritma Porter untuk Bahasa Indonesia dikembangkan oleh Fadillah Z Tala (Tala, 2003) yang mengadopsi Algoritma Porter untuk Bahasa Inggris yang dikembangkan oleh W.B Frakes. Stemmer ini menggunakan rule base analisis untuk mencari root sebuah kata. Stemmer ini sama sekali tidak menggunakan kamus sebagai acuan seperti stemmer Nazief Adriani, Ahmad, Vega atau Jelita. Langkah-langkah pada algoritma Porter adalah:

- 1) Menghapus partikel (“lah”, “kah”, “tah”, “pun”).
- 2) Menghapus kata ganti (Possesive Pronoun), seperti -ku, -mu, -nya
- 3) Menghapus awalan pertama. Jika tidak ditemukan, maka lanjut ke langkah 4a, dan jika ada maka lanjut ke langkah 4b.

- 4) a) Menghapus awalan kedua, dan dilanjutkan pada langkah ke 5a.
b) Menghapus akhiran, jika tidak ditemukan maka kata tersebut diasumsikan sebagai kata dasar (root word). Jika ditemukan maka lanjut ke langkah 5b.
- 5) a) Menghapus akhiran dan kata akhir diasumsikan sebagai kata dasar (root word).
b) Menghapus awalan kedua dan kata akhir diasumsikan sebagai kata dasar (root word). Dalam sebuah kata, memungkinkan adanya dua awalan yang saling berurutan.

2.3.10. Jaro-Winkler

Jaro-Winkler Distance adalah algoritma untuk menghitung nilai jarak kedekatan antara dua teks. Menurut (Winkler, 1999), Jaro-Winkler Distance mempunyai 3 komponen dasar: (1) menghitung panjang string atau kata, (2) mencari nomor huruf pada kedua kata, dan (3) mencari transposisi. Semakin tinggi jarak JaroWinkler Distance antara dua teks berarti semakin ada kemiripan. (Prasetyo, 2018) Jaro-Winkler distance merupakan varian dari Jaro distance metrik yaitu sebuah algoritma untuk mengukur kesamaan antara dua string, biasanya algoritma ini digunakan di dalam pendeteksian duplikat. Semakin tinggi Jaro-Winkler distance untuk dua string, semakin mirip dengan string tersebut. Jaro-Winkler distance terbaik dan cocok untuk digunakan dalam perbandingan string singkat seperti nama orang. Skor normalnya seperti 0 menandakan tidak ada kesamaan, dan 1 adalah sama persis. Algoritma Jaro-Winkler distance memiliki kompleksitas waktu quadratic runtime complexity yang sangat efektif pada string

pendek dan dapat bekerja lebih cepat dari algoritma edit distance. Dasar dari algoritma ini memiliki tiga bagian: (Kurniawati, 2014)

- a. Menghitung panjang string,
- b. Menemukan jumlah karakter yang sama di dalam dua string, dan
- c. Menemukan jumlah transposisi.

Pada algoritma Jaro digunakan rumus untuk menghitung jarak (d_j) antara dua string yaitu s_1 dan s_2 adalah:

$$d_j = \frac{1}{3} \times \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right)$$

dimana :

m = jumlah karakter yang sama persis

$|s_1|$ = panjang string 1

$|s_2|$ = panjang String 2

t = jumlah transposisi

Jarak teoritis dua buah karakter yang disamakan dapat dibenarkan jika tidak melebihi:

$$\left(\frac{\max(|s_1|, |s_2|)}{s} \right) < -1$$

Akan tetapi bila mengacu kepada nilai yang akan dihasilkan oleh algoritma Jaro-Winkler maka nilai jarak maksimalnya adalah 1 yang menandakan kesamaan string yang dibandingkan mencapai seratus persen atau sama persis. Biasanya s_1 digunakan sebagai acuan untuk urutan di dalam mencari transposisi. Yang dimaksud transposisi di sini adalah karakter yang sama dari string yang

dibandingkan akan tetapi tertukar urutannya. Sebagai contoh, dalam membandingkan kata CRATE dengan TRACE, bila dilihat seksama maka dapat dikatakan semua karakter yang ada di s_1 ada dan sama dengan karakter yang ada di s_2 , tetapi dengan urutan yang berbeda. Dengan mengganti C dan T, dapat dilihat perubahan kata CRATE menjadi TRACE. Pertukaran dua elemen string inilah adalah contoh nyata dari transposisi yang dijelaskan. Dalam pencocokkan DwAyNE dan DuANE memiliki urutan yang sama D-A-N-E, jadi tidak ada transposisi.

Jaro-Winkler distance menggunakan prefix scale (p) yang memberikan tingkat penilaian yang lebih, dan prefix length (l) yang menyatakan panjang awalan yaitu panjang karakter yang sama dari string yang dibandingkan sampai ditemukannya ketidaksamaan. Bila string s_1 dan s_2 yang diperbandingkan, maka Jaro-Winkler distancenya (d_w) adalah:

$$d_w = d_j + (lp(1 - d_j)) (3)$$

dimana :

d_j = Jaro distance untuk strings s_1

dan s_2

l = panjang prefiks umum di awal string nilai maksimalnya 4 karakter (panjang karakter yg sama sebelum ditemukan ketidaksamaan max 4)

p = konstanta scaling factor. Nilai standar untuk konstanta ini menurut Winkler adalah $p = 0,1$.

Berikut ini adalah contoh pada perhitungan JaroWinkler distance. Jika string s_1 MARTHA dan s_2 MARHTA maka:

$$m = 6$$

$$s1 = 6$$

$$s2 = 6$$

Karakter yang tertukar hanyalah T dan H. Maka $t = 1$. Maka nilai Jaro distance adalah:

$$d_j = \frac{1}{3} \times \left(\frac{6}{6} + \frac{6}{6} + \frac{6-1}{6} \right) = 0.944$$

Kemudian bila diperhatikan susunan $s1$ dan $s2$ dapat diketahui nilai $l = 3$, dan dengan nilai konstan $p = 0.1$. Maka nilai Jaro-Winkler distance adalah:

$$dw = 0.944 + (3 \times 0.1 (1 - 0.944)) = 0.961$$

Jika string $s1$ DWAYNE dan $s2$ DUANE maka:

$$m = 4$$

$$s1 = 6$$

$$s2 = 5$$

$t = 0$, hal ini dikarenakan tidak ada karakter yang sama tapi tertukar urutannya. Karakter seperti D, A, N, E dianggap dalam urutan yang sama. Maka nilai Jaro distance adalah:

$$d_j = \frac{1}{3} \times \left(\frac{4}{6} + \frac{6}{5} + \frac{4-1}{4} \right) = 0.822$$

Kemudian bila diperhatikan susunan $s1$ dan $s2$ dapat diketahui nilai $l = 1$, dan dengan nilai konstan $p = 0.1$. Maka nilai Jaro-Winkler distance adalah:

$$dw = 0.822 + (1 \times 0.1 (1 - 0.822)) = 0.961$$

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat dan Pendekatan Penelitian

3.1.1. Jenis Penelitian

Jenis penelitian adalah penelitian komputasi eksperimental. Eksperimen yang dilakukan adalah dengan kombinasi algoritma Nazief-Adriani dan Porter Stemmer untuk mengetahui apakah kombinasi algoritma ini memiliki akurasi yang cukup tinggi dalam melakukan stemming bahasa Sasak.

3.1.2. Sifat Penelitian

Dilihat dari segi sifatnya penelitian ini adalah penelitian bersifat deskriptif dengan pemahaman masalah yang baik dan data yang terstruktur.

3.1.3. Pendekatan Penelitian

Pada penelitian ini menggunakan pendekatan penelitian kuantitatif dimana data yang digunakan adalah data yang diambil dari Tesaurus Bahasa Sasak. Data dari buku Tesaurus Bahasa Sasak digunakan sebagai data kamus pada penelitian ini, dengan data sebanyak 2.532 kata.

3.2. Metode Pengumpulan Data

Dalam penelitian ini, pengumpulan data yang akan digunakan menggunakan beberapa langkah yang berkaitan dengan metode penelitian tersebut yaitu dengan metode observasi dan studi literatur atau kepustakaan

3.2.1. Studi literatur atau kepustakaan

Peneliti melakukan pencarian literatur dan pengumpulan materi yang berkaitan dari buku, jurnal dan referensi lainnya yang yang terkait dengan bahasa Sasak. Sehingga diperoleh literatur yang sesuai dengan obyek penelitian.

3.2.2. Observasi

Observasi adalah metode pengumpulan data melalui pengamatan yang cermat dan teliti secara langsung terhadap gejala-gejala yang diselidiki. Observasi yang digunakan adalah observasi tidak langsung, yaitu untuk memperoleh data dari subjek penelitian dengan mengambil data yang akan ditesmming dari social media. Peneliti mengambil data set yang ada kemudian mengolah datanya untuk kemudian dijadikan subjek penelitian (Sarwono, 2006).

3.3. Metode Analisis Data

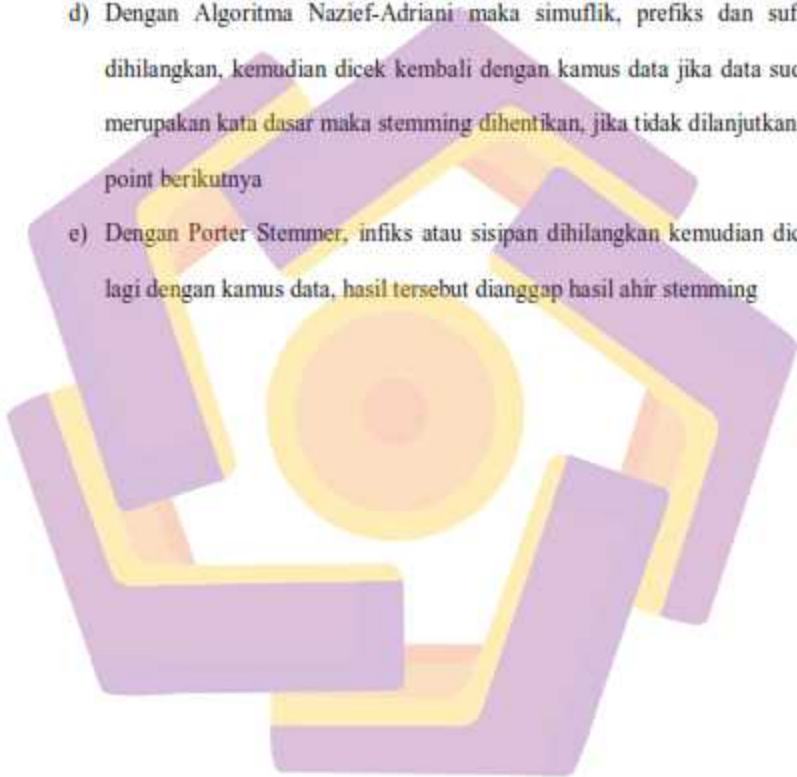
Analisis data dilakukan setelah proses evaluasi selesai. Dari proses evaluasi maka didapatkan jumlah kata yang berhasil di-stemming dan yang gagal. Dari data jumlah kata tersebut kemudian diukur berapa tingkat akurasi yang dapat dihasilkan dari kombinasi algoritma Nazief-Adriani dan Porter Stemmer.

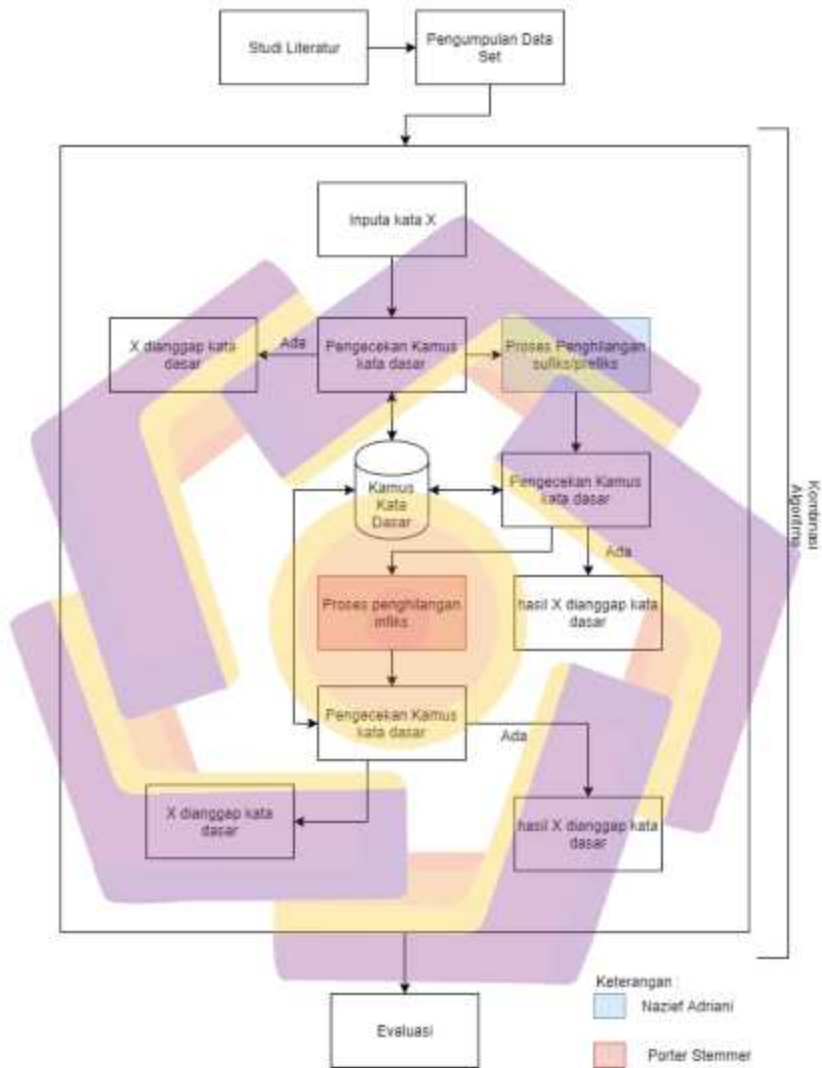
3.4. Alur Penelitian

Alur penelitian dapat dilihat pada Gambar 1, dengan keterangan sebagai berikut:

- a) Pengumpulan data dilakukan dengan input kata dasar dari buku Tesaurus Bahasa Sasak (2016)

- b) kata dinormalisasi dengan cara menyimpan data dengan Array of String supaya dapat diolah oleh aplikasi
- c) Kata dimasukkan dan dicek terlebih dahulu apakah data tersebut sudah merupakan kata dasar, jika bukan kata dasar maka lanjut ke point d)
- d) Dengan Algoritma Nazief-Adriani maka simuflik, prefiks dan sufiks dihilangkan, kemudian dicek kembali dengan kamus data jika data sudah merupakan kata dasar maka stemming dihentikan, jika tidak dilanjutkan ke point berikutnya
- e) Dengan Porter Stemmer, infiks atau sisipan dihilangkan kemudian dicek lagi dengan kamus data, hasil tersebut dianggap hasil akhir stemming





Gambar 3.1. Gambar Ahur Penelitian

Studi literature merupakan proses pembelajaran mendalam bahasa sasak, terutama pada pola perubahan kata seperti imbuhan-imbuhan yang ada pada bahasa sasak. Dalam proses studi literature juga memperdalam teori algoritma nazief-adriani, porter stemmer dan jaro winkler distance.

Pengumpulan data merupakan proses pengambilan data yang nantinya akan menjadi kamus data bahasa Sasak. Data yang didapatkan merupakan himpunan kata dasar bahasa sasak dari buku Tesaurus Bahasa Sasak (Nuryati, 2016). Data diinput secara manual melalui scanner Google Translation kemudian disimpan sebagai file txt.

Kemudian untuk proses kombinasi algoritma sendiri adalah proses stemming bahasa Sasak menggunakan kombinasi algoritma Nadzief-Adriani, Porter Stemmer dan Jaro-Winkler Distance. Terdapat beberapa kemungkinan scenario yang dijalankan:

- a) Kata yang dimasukkan merupakan kata dasar

Terdapat skenario dimana kata yang dimasukan adalah kata dasar. Pada Gambar 1, sebelum menjalankan algoritma maka akan dicek dulu apakah kata yg dimasukkan sudah ada di kamus dasar kata. Jika kata tersebut ada di kamus dasar kata atau ketika dilakukan pengecekan similarity dengan jarak lebih dari 0.9 berarti stemming tidak akan dilakukan. contoh kata yang dimasukkan adalah lalo yang berarti pergi. kata *lalo* tidak memiliki imbuhan sama sekali, dengan begitu lalo adalah kata dasar.

- b) Kata yang dimasukkan memiliki imbuhan tanpa sisipan

Jika kata yang dimasukkan memiliki imbuhan atau bukan merupakan kata dasar maka hasil pencarian di kamus dasar kata bernilai *false*. Algoritma akan melanjutkan untuk melakukan proses *Enhanced Confix Stripping*. Setelah diketahui urutan penghilangan antara sufiks dan prefiks kata akan diproses menggunakan algoritma Nazief-Adriani. Setelah hasil penghilangan prefiks/sufiks didapatkan maka akan dicek lagi menggunakan kamus kata dasar. Jika hasil stemming ada di kamus kata dasar maka proses dihentikan. Contoh menggunakan kata *tetaletan*, kata tersebut memiliki kata dasar *talet* yang berarti tanam dan imbuhan *te-an* yang mengubah artinya menjadi tanaman.

c) Kata yang dimasukkan memiliki imbuhan dengan sisipan

Jika kata yang dimasukkan memiliki sisipan, maka proses yang dilakukan sama seperti point b) tetapi hasil pencarian di kamus kata untuk kedua kalinya tentunya akan bernilai *false*. Karena masih bernilai *false*, maka akan dilanjutkan stemming menggunakan algoritma Porter Stemmer. Terlepas dari ada atau tidaknya hasil dari algoritma tersebut terdapat di kamus kata dasar, point ini memberikan hasil akhir. Jika hasil stemming memang ada di kamus kata dasar maka stemming berhasil, jika tidak ada kemungkinan terdapat overstemming atau understemming tetapi hasil ini akan tetap ditampilkan. Contoh menggunakan kata *begegoloqan* yang memiliki arti tidur-tiduran, *be-an* merupakan simufleks (gabungan awalan dan akhiran) yang kemudian menyisakan *gegoloq* dengan sisipan *-eg-*. Kata dasar dari *begegoloqan* adalah *goloq*.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Pengumpulan Dan Normalisasi Data

Data yang didapatkan merupakan himpunan kata dasar bahasa sasak dari buku Tesaurus Bahasa Sasak (Nuryati, 2016). Buku tersebut berisi 2.532 kata dasar bahasa sasak beserta penjelasannya dalam bahasa sasak serta keterangan kata tersebut merupakan kata benda, kata kerja atau kata keterangan. Dari isi buku tersebut yang diambil adalah kata dasar saja tanpa mengambil data keterangan kata.

Tabel 4.1. Hasil kumpulan kata dasar bahasa sasak

| No | Kata Dasar (Sasak) |
|----|--------------------|
| 1 | ande |
| 2 | andok |
| 3 | aneh |
| 4 | angen |
| 5 | angkaq |
| 6 | angkun |
| 7 | anjar |
| 8 | antap |
| 9 | antem |
| 10 | antep |
| 11 | antih |
| 12 | anyon |
| 13 | anyong |
| 14 | aoq |
| 15 | ape |
| 16 | apeq |
| 17 | apokat |
| 18 | apuh |
| 19 | aran |
| 20 | araq |
| 21 | ariq |
| 22 | aroh |

4.2. Modifikasi Algoritma

Algoritma dimodifikasi dan diimplementasikan menggunakan bahasa pemrograman Java. Modifikasi dilakukan pada source code pihak ketiga untuk stemming kata menggunakan algoritma Nazief-Adriani dari github (<https://github.com/arumsha/stemmingNaziefAndriani>). Source code ini spesifik untuk stemming kata dalam Bahasa Indonesia maka modifikasi dilakukan untuk penyesuaian imbuhan dalam Bahasa Sasak.

Seperti penjelasan yang dibahas pada landasan teori Nazief-Adriani, dijelaskan bahwa algoritma Nazief-Adriani menghilangkan Inflectional suffixes, yaitu dengan menghilangkan particle (“-lah”, “-kah”, “-tah” atau “-pun”), kemudian hilangkan inflectional possessive pronoun suffixes (“-ku”, “-mu” atau “-nya”) pada Bahasa Indonesia.

Dalam modifikasi algoritma Nazief-Adriani untuk Bahasa Sasak menghapus akhiran “-an”, “-ang”, “-n”, “-m”, “-ne” dan “-ku”. Akhiran “-ku” bisa berarti 2 posisi yaitu -ku dalam arti kepemilikan dan -ku sebagai pengganti subjek. Karena dalam Bahasa Sasak subjek yang bergabung dengan kata kerja ditempatkan sebagai akhiran, bukan awalan. Jika diibaratkan dengan Bahasa Indonesia yang berarti “saya pergi”, dalam Bahasa Indonesia akan ditulis “Ku pergi” maka dalam Bahasa Sasak ditulis “laloku” dimana *lalo* berarti pergi dan -ku disini sebagai subjek. Walaupun dalam Bahasa Indonesia bisa juga ditulis “pergiku” tapi kata ini berarti “kepergianku”, sedangkan dalam Bahasa Sasak artinya tetap “saya pergi”.

Selain akhiran, awalan juga dibersihkan. Awalan dalam Bahasa Sasak tak jauh berbeda dari Bahasa Indonesia. Awalan yang dibersihkan adalah “be-“,

“ber”, “pe-”, “pe-”, “peng-”, “me-” dan “nge-”. Tetapi sebelum melakukan *stem* untuk awalan atau akhiran, yang paling pertama dilakukan adalah *stem* untuk simufliks yaitu imbuhan kombinasi awalan dan akhiran. Hal ini dilakukan untuk mengantisipasi terjadinya *overstem* pada awalan dan akhiran. Berikut adalah simufliks dalam Bahasa Sasak; “be-an”, “pe-an”, “ke-an” dan “te-an”.

Modifikasi lain terdapat pada source code pihak ketiga untuk porter stemmer (<http://www.tartarus.org/~martin/PorterStemmer>). Source code tersebut spesifik mendeteksi imbuhan untuk kata dalam Bahasa Inggris, dan untuk Porter Stemmer yang dimofikasi untuk Bahasa Indonesia sudah dilakukan oleh Fadillah Z Taala (Tala, 2002), maka modifikasi dilakukan untuk penyesuaian Bahasa Sasak, dan khusus untuk mendeteksi sisipan kata karena awalan dan akhiran sudah dilakukan dengan algoritma Nazief-Adriani. Berikut modifikasi langkah yang dilakukan Taala untuk stemmer Bahasa Indonesia:

- 1) Menghapus partikel (“lah”, “kah”, “tah”, “pun”).
- 2) Mengapus kata ganti (Possesive Pronoun), seperti -ku, -mu, -nya
- 3) Mengapus awalan pertama. Jika tidak ditemukan, maka lanjut ke langkah 4a, dan jika ada maka lanjut ke langkah 4b.
- 4) a) Menghapus awalan kedua, dan dilanjutkan pada langkah ke 5a.
b) Menghapus akhiran, jika tidak ditemukan maka kata tersebut diasumsikan sebagai kata dasar (root word). Jika ditemukan maka lanjut ke langkah 5b.
- 5) a) Menghapus akhiran dan kata akhir diasumsikan sebagai kata dasar (root word).

b) Menghapus awalan kedua dan kata akhir diasumsikan sebagai kata dasar (root word). Dalam sebuah kata, memungkinkan adanya dua awalan yang saling berurutan.

Kemudian untuk Bahasa Sasak imbuhan yang dicari disesuaikan untuk Bahasa Sasak seperti imbuhan yang diterapkan pada modifikasi Nazief-Adriani. Tambahan untuk Porter Stemmer adalah mendeteksi sisipan dalam Bahasa Sasak yaitu; “-er-”, “-el-”, “-eg-”, dan “-em-”.

Untuk pengecekan similarity tidak dilakukan modifikasi, source code diambil dari web official apache dengan sama-sama menggunakan bahasa pemrograman java (<https://commons.apache.org/sandbox/commons-text/jacoco/org.apache.commons.text.similarity/JaroWinklerDistance.java.html>).

Setelah masing-masing algoritma diimplementasikan, hasil stemming dicetak dalam bentuk file excel untuk mengetahui seberapa banyak tingkat ketepatan atau kesalahan yang dihasilkan.

4.3. Hasil Stemming

Algoritma ini merupakan hasil kombinasi dari algoritma Nazief-Adriani dan Porter Stemmer dengan pengecekan kamus data langsung menggunakan kata dan pengecekan similarity menggunakan algoritma Jaro-Winkler. Penerapan similarity dapat dilihat pada gambar 2.



Gambar 4.1. Pseudocode penerapan similarity

Penerapan similarity ini tidak mengambil jarak mutlak 1 tetapi mengeset *default* jarak awal sebesar 0.95 kemudian kata inputan dibandingkan dengan semua kata yang ada dalam kamus kata. Kata yang memiliki jarak paling tinggi (yang melebihi jarak *default*) dianggap sebagai kata dasarnya.

Hasil stemming dengan kombinasi Nadzief-Adriani dan Porter Stemmer beserta pengecekan kamus kata menggunakan similarity algoritma Jaro-Winkler terdapat pada Table 4.2.

Tabel 4.2. Hasil Stemming Kombinasi Nazief-Adriani dan Porter Stemmer
Beserta Jaro-Winkler Distance (Lanjut di Lampiran)

| <i>Word</i> | Found on dictionary | Stem word | Found by similarity | Similarity distance | Similar word | SUCCESS |
|------------------|----------------------------|------------------|----------------------------|----------------------------|---------------------|----------------|
| <i>leq</i> | false | leq | true | 0.93 | laeq | 0 |
| <i>sopoq</i> | true | sopoq | false | 0.0 | | 1 |
| <i>jelo</i> | false | jelo | true | 0.96 | jeloh | 0 |
| <i>dua</i> | false | dua | false | 0.00 | dugaq | 0 |
| <i>kanak</i> | true | kanak | false | 0.0 | | 1 |
| <i>siq</i> | true | siq | false | 0.0 | | 1 |
| <i>besemeton</i> | false | semeton | true | 1.0 | semeton | 1 |

Dari hasil stemming dengan kombinasi algoritma Nazief-Adriani dan Porter Stemmer serta pengecekan similarity menggunakan Jaro-Winkler Distance pada Table 4.2 menunjukkan bahwa algoritma ini menghasilkan 859 kata yang berhasil di-stemming dan 531 kata yang gagal di-stemming dari 2571 kata yang diinputkan yang akhirnya di-distinc menjadi 1390. Hasil pengecekan similarity stemming menggunakan Jaro-Winkler Distance masih menunjukkan beberapa kata yang tidak sesuai dengan kata dasar yang diharapkan dikarenakan ada beberapa kata yang memiliki jarak paling dekat (misal 0.98) misal dalam kamus data ada *anaq* dan *araq*. Walaupun dengan jarak yang sangat dekat kata tersebut bukanlah hasil stemming yang sesuai.

Tabel 4.3. Hasil Stemming Kombinasi Nazief-Adriani dan Porter Stemmer
Tanpa Jaro-Winkler Distance (Lanjutan di Lampiran)

| <i>Word</i> | Found on dictionary | Stem word | Found by similarity | Similarity distance | SUCCESS |
|--------------|----------------------------|------------------|----------------------------|----------------------------|----------------|
| <i>leq</i> | false | leq | false | 0.0 | 1 |
| <i>sopoq</i> | true | sopoq | false | 0.0 | 1 |
| <i>jelo</i> | false | jelo | false | 0.0 | 1 |
| <i>dua</i> | false | dua | false | 0.0 | 1 |
| <i>kanak</i> | true | kanak | false | 0.0 | 1 |

| <i>Word</i> | Found on dictionary | Stem word | Found by similarity | Similarity distance | SUCCESS |
|------------------|----------------------------|------------------|----------------------------|----------------------------|----------------|
| <i>siq</i> | true | siq | false | 0.0 | 1 |
| <i>beseleton</i> | false | seleton | false | 0.0 | 1 |

Dari hasil stemming dengan kombinasi algoritma Nazief-Adriani dan Porter Stemmer tanpa pengecekan similarity pada table 4.3 menunjukkan bahwa algoritma ini menghasilkan 861 kata yang berhasil di-stemming dan 529 kata yang gagal di-stemming dari 2571 kata yang diinputkan yang akhirnya di-distinc menjadi 1390. Hasil tanpa similarity check menunjukkan kenaikan akurasi sebanyak 0,359%.

Tabel 4.5. Hasil Stemming Kombinasi Nazief-Adriani dan Porter Stemmer Tanpa Jaro-Winkler Distance dan Tanpa Dictionary (Lanjut di Lampiran)

| <i>Word</i> | Found on dictionary | Stem word | SUCCESS |
|------------------|----------------------------|------------------|----------------|
| <i>leq</i> | false | leq | 1 |
| <i>sopoq</i> | false | sopoq | 1 |
| <i>jelo</i> | false | jelo | 1 |
| <i>dua</i> | false | dua | 1 |
| <i>kanak</i> | false | kanak | 1 |
| <i>siq</i> | false | siq | 1 |
| <i>beseleton</i> | false | seleton | 1 |
| <i>sino</i> | false | sino | 1 |
| <i>tekocapan</i> | false | tekocapan | 0 |
| <i>kelapahan</i> | false | kelapahan | 0 |
| <i>lengan</i> | false | lengan | 1 |
| <i>jelo</i> | false | jelo | 1 |
| <i>siq</i> | false | siq | 1 |
| <i>rubin</i> | false | rubin | 1 |
| <i>laiq</i> | false | laiq | 1 |
| <i>ndeqna</i> | false | ndeqna | 0 |

Dari hasil stemming dengan kombinasi algoritma Nazief-Adriani dan Porter Stemmer tanpa pengecekan similarity dan tanpa dictionary pada Table 4.5.

menunjukkan bahwa algoritma ini menghasilkan 690 kata yang berhasil di-*stemming* dan 700 kata yang gagal di-*stemming* dari 2571 kata yang diinputkan yang akhirnya di-*distinc* menjadi 1390. Adanya kamus data sangat berpengaruh pada hasil stemming.

Tabel 4.6. Hasil Stemming Nazief-Adriani Tanpa Jaro-Winkler Distance

| Word | Found on dictionary | Stem word | SUCCESS |
|------------------|----------------------------|------------------|----------------|
| <i>leq</i> | false | leq | 1 |
| <i>sopoq</i> | true | sopoq | 1 |
| <i>jelo</i> | false | jelo | 1 |
| <i>dua</i> | false | dua | 1 |
| <i>kanak</i> | true | kanak | 1 |
| <i>siq</i> | true | siq | 1 |
| <i>besemeton</i> | false | besemeton | 0 |
| <i>sino</i> | false | sino | 1 |
| <i>tekocapan</i> | false | tekocapan | 0 |
| <i>kelapahan</i> | false | kelapahan | 0 |
| <i>lengan</i> | false | lengan | 1 |
| <i>jelo</i> | false | jelo | 1 |
| <i>siq</i> | true | siq | 1 |
| <i>rubin</i> | true | rubin | 1 |
| <i>laiq</i> | true | laiq | 1 |
| <i>ndeqna</i> | false | ndeqna | 1 |

Dari hasil stemming Nazief-Adriani tanpa pengecekan similarity pada table 4.6 menunjukkan bahwa algoritma ini menghasilkan 685 kata yang berhasil di-*stemming* dan 705 kata yang gagal di-*stemming* dari 2571 kata yang diinputkan yang akhirnya di-*distinc* menjadi 1390. Perbedaan ada tidaknya stemmer tidak terlihat signifikan karena bahasa sasak lebih banyak menggunakan kata yang sederhana.

Masing-masing kombinasi algoritma memiliki persentase yang berbeda-beda dalam melakukan stemming Bahasa Sasak. Perbandingan tingkat akurasi yang dihitung berdasarkan persentase benar dan salah yang mengacu pada table

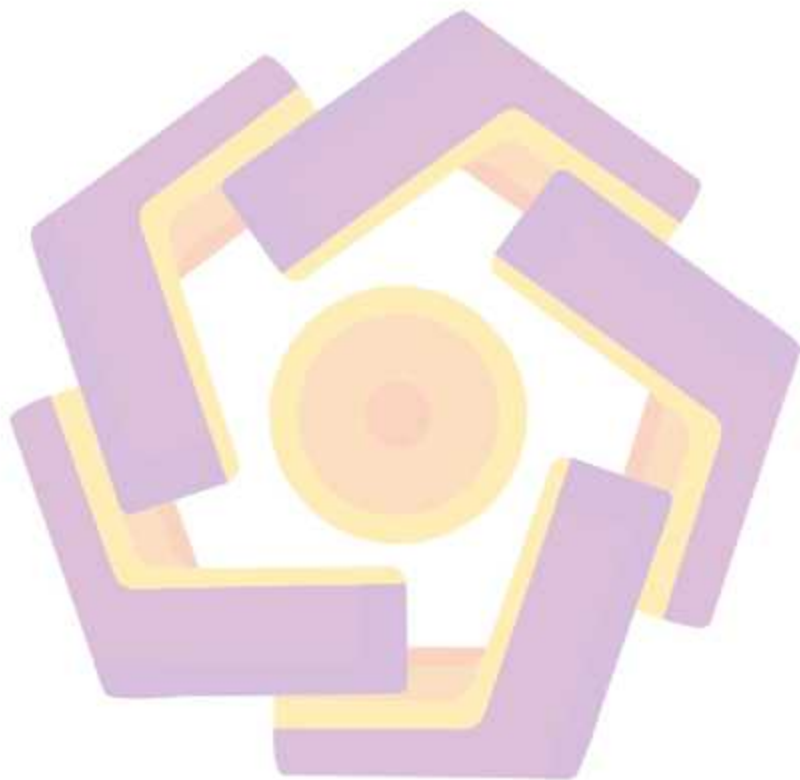
4.2, 4.3, 4.5 dan 4.6 dari masing-masing kombinasi algoritma dapat dilihat pada Table 4.7.

Tabel 4.7. Perbandingan Tingkat keberhasilan *stemming* kata pada Bahasa Sasak

| Nama Kombinasi | Kombinasi Nazief-Adriani dan Porter Stemmer Beserta Jaro-Winkler Distance | Kombinasi Nazief-Adriani dan Porter Stemmer Tanpa Jaro-Winkler Distance | Kombinasi Nazief-Adriani dan Porter Stemmer Tanpa Jaro-Winkler Distance dan Tanpa Dictionary | Nazief-Adriani Tanpa Jaro-Winkler Distance |
|-----------------------|--|--|---|---|
| Benar | 859 | 861 | 690 | 685 |
| Salah | 531 | 529 | 700 | 705 |
| Akurasi | 61,790% | 61,942% | 49,640% | 49,280% |

Dari Table 4.7. dapat dilihat bahwa presentase keberhasilan menggunakan kombinasi Nazief-Adriani, Porter Stemmer dan Jaro Winkler memiliki akurasi sebanyak 61,790%. Sedangkan tanpa Jaro-Winkler menghasilkan 61,942% hal ini diakibatkan adanya overstem yang dilakukan oleh Jaro-winkler. Kemudian dalam kobinasi-kobinasi diatas jika dictionary atau kamus kata dihilangkan akurasi

menurun menjadi 49,640%. Hal ini membuktikan kamus data dapat menambah akurasi sebanyak 1,23%.



BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil evaluasi dan analisa yang telah dilakukan maka dapat ditarik kesimpulan sebagai berikut:

1. Berdasarkan hasil evaluasi kombinasi Nazief-Adriani dan Porter Stemmer menggunakan kamus data dengan 2.532 kata dasar dan menggunakan *similarity check* dengan algoritma Jaro-Winkler disimpulkan bahwa kombinasi ini dapat digunakan untuk stemming bahasa Sasak dengan tingkat keberhasilan 61,790% dengan sample data yang diolah untuk *stemming* sebanyak 2.571 kata yang akhirnya di-*distinc* menjadi 1390 dan berhasil stemming sebanyak 859 kata. Nilai keberhasilan ini dianggap cukup karena berhasil stemming setengah dari sample data yang disediakan tetapi masih memerlukan improvisasi.
2. Penerapan kamus data dapat berperan dan mempengaruhi persentasi keberhasilan stemming. Dengan kamus data dengan 2.532 kata dasar menambahkan kesuksesan stemming sebanyak 1,23% dengan sample data yang diolah untuk *stemming* sebanyak 2.571 kata yang akhirnya di-*distinc* menjadi 1390 kata, kemudian dibandingkan stemming menggunakan kamus data dan tanpa

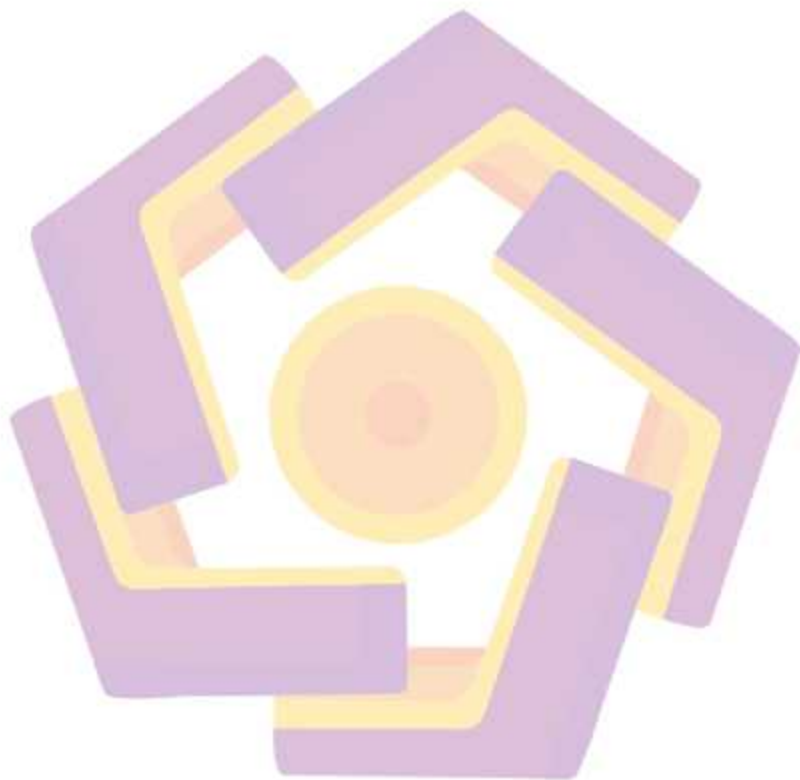
kamus data terdapat perbedaan sebanyak 171 data. 171 data tersebut menjadi keunggulan diterapkannya kamus data.

3. Dengan adanya similarity check menggunakan Jaro-Winkler algorithm dengan default similarity point sebanyak 0.95 (skala dari 0 sampai 1) menambah overstem sebanyak 0,359% atau dengan sample data yang diolah untuk *stemming* sebanyak 2.571 kata yang akhirnya di-*distine* menjadi 1390 kata, kemudian dibandingkan stemming menggunakan similarity check dan tanpa similarity check terdapat perbedaan sebanyak 5 data. 5 data tersebut menjadi kelemahan diterapkannya similarity check.
4. Dari 4 dialek yang ada dalam Bahasa Sasak yaitu Dialek Bayan, Dialek Pujut, Dialek Aiq Bukaq dan Dialek Selaparang, penelitian ini fokus pada dialek Pujut.
5. Manfaat teoritis dan praktis dari penelitian ini telah dicapai karena menghasilkan kata dasar yang diolah melalui stemming dengan kombinasi Nazief-Adriani dan Porter Stemmer.

5.2 Saran

1. Penelitian selanjutnya dapat memperhatikan perbedaan dan keragaman dialek bahasa Sasak yang mempengaruhi penulisan dan pengucapan.
2. Memperkaya kamus data dasar agar hasil stemming dapat ditingkatkan.
3. Menggunakan contoh yang lebih beragam.
4. Dengan adanya overstem karena similarity chek menggunakan algoritma Jaro-Winkler dapat ditingkatkan dengan menambahkan kamus data yang lebih lengkap atau meeningikan default similarity point.

5. Hasil stemming dari penelitian ini dapat digunakan untuk sentiment analysis untuk penelitian lebih lanjut.



DAFTAR PUSTAKA

Pustaka Buku

Bird, S. (2014). *Natural Language Processor*. O'REILLY.

Hakim, L. (2016). *Ensiklopedia Bahasa Sasak*. Mataram: Kantor Bahasa Nusa Tenggara Barat.

Nuryati. (2016). *Tesaurus Bahasa Sasak*. Mataram: Kantor Bahasa Nusa Tenggara Barat.

Sarwono. (2006). *Metode Penelitian Kuantitatif dan Kualitatif*. Yogyakarta: Graha Ilmu.

Pustaka Majalah, Jurnal Ilmiah atau Prosiding

Asian, J. (2007). Effective Techniques for Indonesian Text Retrieval. *School of Computer Science and Information Technology, Science, Engineering, and Technology Portfolio, RMIT University, Melbourne, Victoria, Australia*.

Balasanakar. (2016). Multi Level Inflection Handling Stemmer using Iterative Suffix Stripping for Malayalam Language. *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp.530-534.

Kaara, W. B. (2013). A New Stemmer to improve Information Retrieval. *International Journal of Network Security & Its Application (IJNSA)*.

- Maylawati, D. S. (2018). An Improved of Stemming Algorithm for Mining Indonesian Text with Slang on Social Media. *The 6th International Conference on Cyber and IT Service Management (CITSM 2018)*.
- Nata, G. N. (2017). Stemming Teks Sor-Singgih Bahasa Bali. *Konferensi Nasional Sistem & Informatika 2017*.
- Patil. (2017). MarS: A RuleBased Stemmer for Morphologically Rich Language Marathi. *International Conference on Computer, Communications and Electronics*. pp.580-584.
- Pramudita. (2018). Klasifikasi Berita Olahraga menggunakan Metode Naive Bayes dengan Enhanced Confix Stripping Stemmer. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol.5, no.3, pp.269-276.
- Putra, R. B. (2018). Non-formal Affixed Word Stemming in Indonesian Language. *International Conference on Information and Communications Technology (ICOIACT)*.
- Subali, M. A. (2019). Kombinasi Metode Rule-Based Dan N-Gram Stemming Untuk Mengenali Stemmer Bahasa Bali. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) Vol. 6, No. 2, April 2019*.
- Widayanto, H. (2017). Comparison Nazief Adriani And CS Stemmer Algorithm For Stemm Real Data. *e-Proceeding of Engineering : Vol.4, No.3 Desember 2017*.
- Widjaja, M. (2015). Implementation of Porter's Modified Stemming Algorithm in an Indonesian Word Error Detection Plugin Application. *International Journal of Technology (2015) 2: 139-150*.

Wiliam, S. (2006). Tingkat Tutur dalam Bahasa Sasak dan Bahasa Jawa.

WACANA VOL. 8 NO. 1, APRIL 2006.

Kurniawati, Anna. (2014). Implementasi Algoritma Jaro-Winkler Distance untuk

Membandingkan Kesamaan Dokumen Berbahasa Indonesia. *Gunadarma,*

2014

Prasetyo Agung. *Algoritma Jaro-Winkler Distance: Fitur Autocorrect Dan*

Spelling Suggestion Pada Penulisan Naskah Bahasa Indonesia di BMS TV.

Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK) DOI:

10.25126/jtik.201854780 Vol. 5, No. 4, September 2018, hlm. 435-444

Pustaka Laporan Penelitian

Tala, F. (2002). A Study of Stemming Effect on Information Retrieval in Bahasa

Indonesia. *Universiteit van Amsterdam.*

Baiq, Justiana Afianty. (2013). Analisis Kontrastif Sintaksis Bahasa Indonesia

Dan Bahasa Sasak Dialek {A-E} di Kelurahan Dasan Agung Kota

Mataram Serta Implikasinya Terhadap Pembelajaran Menulis Di Smp.

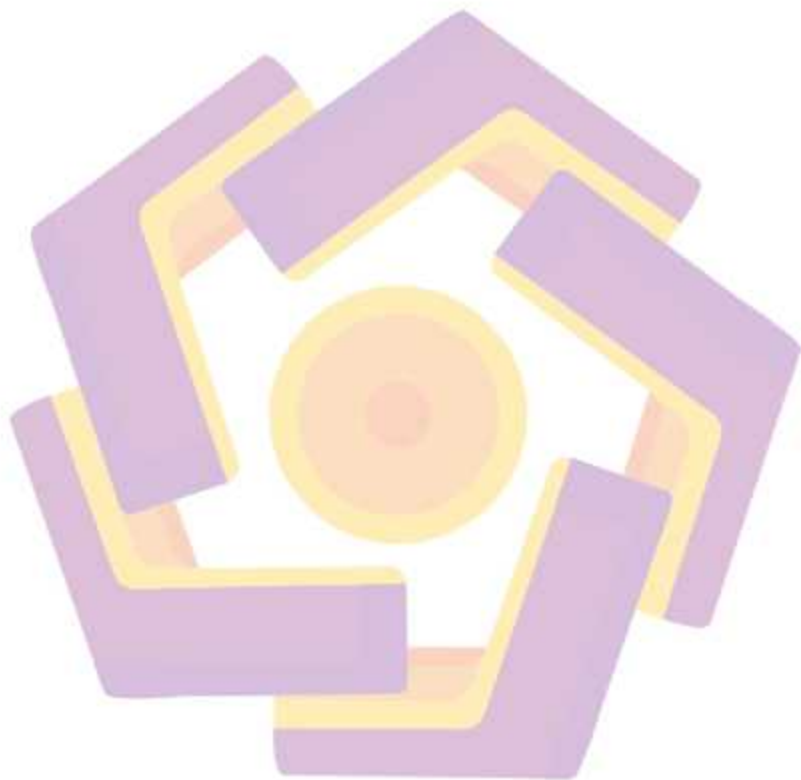
Universiteit Mataram.

Pustaka Elektronik

Kemendikbud, "Bahasa dan Peta Bahasa di Indonesia." [Online].

Available: <http://petabahasa.kemdikbud.go.id>

LAMPIRAN



Lampiran Tabel 5.1 Hasil Normalisasi Data

Tabel 5.1 Hasil Normalisasi Data

| No | Kata | No | Kata | No | Kata | No | Kata |
|----|--------|-----|---------|------|-----------|------|-----------|
| 1 | abah | 634 | bukit | 1267 | judi | 1900 | payas |
| 2 | abang | 635 | bukti | 1268 | jujuh | 1901 | payu |
| 3 | abon | 636 | bukuq | 1269 | jujul | 1902 | pecat |
| 4 | abot | 637 | bukus | 1270 | jujur | 1903 | pecok |
| 5 | abut | 638 | bulat | 1271 | jukung | 1904 | pedaq |
| 6 | aceh | 639 | bulet | 1272 | julat | 1905 | pedas |
| 7 | acong | 640 | buncut | 1273 | julu | 1906 | pejen |
| 8 | adal | 641 | bungkah | 1274 | juluk | 1907 | pejet |
| 9 | adam | 642 | bungkak | 1275 | jungaq | 1908 | pekak |
| 10 | adang | 643 | bungkus | 1276 | jungkat | 1909 | pekek |
| 11 | adat | 644 | bungul | 1277 | juntang | 1910 | pekel |
| 12 | adek | 645 | bungus | 1278 | jurak | 1911 | pekir |
| 13 | ngadek | 646 | bunter | 1279 | juret | 1912 | pekok |
| 14 | adeng | 647 | buntet | 1280 | juwung | 1913 | pekor |
| 15 | adep | 648 | buntu | 1281 | kabar | 1914 | pelai |
| 16 | adeq | 649 | buntung | 1282 | kabeh | 1915 | pele |
| 17 | adil | 650 | buntut | 1283 | kabul | 1916 | pelemeng |
| 18 | adiq | 651 | bunuh | 1284 | kace | 1917 | peleng |
| 19 | adon | 652 | bur | 1285 | kacep | 1918 | pelesir |
| 20 | adok | 653 | burak | 1286 | kacer | 1919 | pelet |
| 21 | adu | 654 | buras | 1287 | kadang | 1920 | pelet |
| 22 | ngadu | 655 | burik | 1288 | kadaq | 1921 | pelewas |
| 23 | aduk | 656 | burung | 1289 | kadi | 1922 | pelih |
| 24 | aer | 657 | busuk | 1290 | kado | 1923 | peling |
| 25 | agem | 658 | butak | 1291 | kadu | 1924 | pelinggih |
| 26 | aget | 659 | buteng | 1292 | kaeh | 1925 | pelit |
| 27 | agiq | 660 | butes | 1293 | kaing | 1926 | pelot |
| 28 | agol | 661 | butir | 1294 | kaji | 1927 | pelungguh |
| 29 | agum | 662 | butoq | 1295 | kakak | 1928 | pempang |
| 30 | agung | 663 | buyut | 1296 | nengkakak | 1929 | penarak |
| 31 | ahat | 664 | cacaq | 1297 | kakaq | 1930 | pendaq |
| 32 | aherat | 665 | cacar | 1298 | kakas | 1931 | penenteng |
| 33 | ahir | 666 | cacat | 1299 | pekakas | 1932 | pengah |
| 34 | ahlak | 667 | cace | 1300 | kaken | 1933 | pengas |
| 35 | ahli | 668 | cacun | 1301 | kakoq | 1934 | pengget |
| 36 | aing | 669 | cakalan | 1302 | kakus | 1935 | pengkung |
| 37 | aip | 670 | cakok | 1303 | kalah | 1936 | pengot |
| 38 | ajah | 671 | calon | 1304 | kalam | 1937 | pengulu |

| | | | | | | | |
|----|-----------|-----|----------|------|----------|------|------------|
| 39 | ajak | 672 | campah | 1305 | kalangan | 1938 | penjare |
| 40 | ajal | 673 | campur | 1306 | kaliq | 1939 | penoq |
| 41 | ajaq | 674 | camut | 1307 | kalis | 1940 | pentang |
| 42 | ajat | 675 | cangkit | 1308 | kalkun | 1941 | peos |
| 43 | ajong | 676 | cantel | 1309 | kalong | 1942 | pepek |
| 44 | ajor | 677 | canting | 1310 | kambaq | 1943 | pepeq |
| 45 | ngajor | 678 | cap | 1311 | kambes | 1944 | pepit |
| 46 | ajum | 679 | caplak | 1312 | kamu | 1945 | perai |
| 47 | akal | 680 | caprek | 1313 | kanak | 1946 | peraq |
| 48 | akar-akar | 681 | capuh | 1314 | kanan | 1947 | perasmanan |
| 49 | akat | 682 | care | 1315 | kancee | 1948 | perau |
| 50 | akbar | 683 | caruk | 1316 | kandoq | 1949 | peredah |
| 51 | akekah | 684 | catet | 1317 | kangroq | 1950 | perekeq |
| 52 | akil | 685 | catut | 1318 | kangkeq | 1951 | peren |
| 53 | akon | 686 | cebet | 1319 | kanji | 1952 | pereq |
| 54 | aku | 687 | cecel | 1320 | kantong | 1953 | peres |
| 55 | akuq | 688 | ceceret | 1321 | kaok | 1954 | peret |
| 56 | alang | 689 | cegrek | 1322 | kaos | 1955 | periak |
| 57 | alat | 690 | cekel | 1323 | kaot | 1956 | periap |
| 58 | ale-ale | 691 | ceket | 1324 | kapak | 1957 | peridug |
| 59 | alep | 692 | cekok | 1325 | kaper | 1958 | perikak |
| 60 | alias | 693 | celang | 1326 | kapok | 1959 | peririq |
| 61 | alih | 694 | cele | 1327 | kapuk | 1960 | periris |
| 62 | alim | 695 | celilong | 1328 | kapur | 1961 | perok |
| 63 | alir | 696 | celingah | 1329 | kaput | 1962 | perot |
| 64 | alit | 697 | celup | 1330 | karat | 1963 | persiq |
| 65 | aloh | 698 | celur | 1331 | karek | 1964 | peru |
| 66 | aluan | 699 | kecelur | 1332 | karing | 1965 | peruq |
| 67 | aluh | 700 | celut | 1333 | kasap | 1966 | perus |
| 68 | alung | 701 | cemoh | 1334 | kasaq | 1967 | perut |
| 69 | aluq | 702 | cemus | 1335 | kasar | 1968 | pesis |
| 70 | alur | 703 | cengak | 1336 | kasoh | 1969 | pesor |
| 71 | alus | 704 | cengen | 1337 | katoq | 1970 | pesot |
| 72 | amal | 705 | cengiq | 1338 | katuq | 1971 | pete |
| 73 | aman | 706 | cengok | 1339 | kaur | 1972 | peteng |
| 74 | amaq | 707 | kecengok | 1340 | nengkaur | 1973 | peter |
| 75 | amar | 708 | centong | 1341 | kawan | 1974 | peti |
| 76 | ambal | 709 | cepeh | 1342 | kawaq | 1975 | petoq |
| 77 | ambar | 710 | cepeq | 1343 | kawih | 1976 | pijet |
| 78 | ambe | 711 | ceper | 1344 | kawin | 1977 | pikir |
| 79 | amben | 712 | ceprio | 1345 | kawiran | 1978 | piluq |
| 80 | ambes | 713 | ceraput | 1346 | kayang | 1979 | pindah |
| 81 | ambot | 714 | cerebik | 1347 | kaye | 1980 | pineng |
| 82 | ambuk | 715 | ceret | 1348 | kayun | 1981 | pinter |

| | | | | | | | |
|-----|-------------|-----|------------|------|-----------|------|----------|
| 83 | ambuq | 716 | ceringang | 1349 | ke | 1982 | pipit |
| 84 | ambur | 717 | cerite | 1350 | kebal | 1983 | piring |
| 85 | ame | 718 | cermen | 1351 | kebawiq | 1984 | pisang |
| 86 | amer | 719 | cerongcong | 1352 | keber | 1985 | pitnah |
| 87 | amet | 720 | cerubak | 1353 | kebus | 1986 | poger |
| 88 | amet | 721 | cet-cet | 1354 | nengkebus | 1987 | polak |
| 89 | samet | 722 | cicil | 1355 | kebut | 1988 | ponggoq |
| 90 | amil | 723 | cikar | 1356 | kebut | 1989 | ponjol |
| 91 | amok | 724 | cingak | 1357 | kecapin | 1990 | ponjoq |
| 92 | ampah | 725 | ciwe | 1358 | kecepon | 1991 | poroq |
| 93 | ampan | 726 | coak | 1359 | keceq | 1992 | poto |
| 94 | ampang | 727 | cobaq | 1360 | keciput | 1993 | prinsang |
| 95 | ngampang | 728 | cocok | 1361 | kecuali | 1994 | pucuk |
| 96 | ampar | 729 | cocor | 1362 | kedawuk | 1995 | puger |
| 97 | ampas | 730 | cokot | 1363 | kedebong | 1996 | pugut |
| 98 | ampen | 731 | colaq | 1364 | kedek | 1997 | pulih |
| 99 | ampes | 732 | colaq | 1365 | kedeng | 1998 | pundut |
| 100 | ampet | 733 | kecolaq | 1366 | kediq | 1999 | punik |
| 101 | ampik-ampik | 734 | colok | 1367 | keduk | 2000 | puntiq |
| 102 | ampiq | 735 | coloq | 1368 | kedung | 2001 | pupaq |
| 103 | ampir | 736 | comon | 1369 | kejaq | 2002 | pupuk |
| 104 | ngampir | 737 | congak | 1370 | kejem | 2003 | puput |
| 105 | ampong | 738 | congang | 1371 | kejong | 2004 | purah |
| 106 | ampoq | 739 | congtag | 1372 | kejot | 2005 | purname |
| 107 | ampun | 740 | congok | 1373 | ngejot | 2006 | puse |
| 108 | ancot | 741 | contlak | 1374 | kekah | 2007 | rabas |
| 109 | andang | 742 | cop | 1375 | kekah | 2008 | raboq |
| 110 | ande | 743 | copet | 1376 | kekel | 2009 | rage |
| 111 | andek | 744 | coreq | 1377 | kekeng | 2010 | rajek |
| 112 | andos | 745 | corong | 1378 | kekes | 2011 | rak |
| 113 | anduh | 746 | cotek | 1379 | kelah | 2012 | rambok |
| 114 | aneh | 747 | cuat | 1380 | kelambi | 2013 | rambu |
| 115 | angen | 748 | cucuk | 1381 | kelambu | 2014 | rampas |
| 116 | anget | 749 | cucuq | 1382 | kelampan | 2015 | rampih |
| 117 | anget | 750 | cukem | 1383 | kelaq | 2016 | rampok |
| 118 | anggar | 751 | cukup | 1384 | kelas | 2017 | ran |
| 119 | anggep | 752 | cukur | 1385 | kelebet | 2018 | randu |
| 120 | anggo | 753 | culuk | 1386 | kelek | 2019 | rang |
| 121 | anggoq | 754 | cundang | 1387 | keleleq | 2020 | rangges |
| 122 | angkan | 755 | dadar | 1388 | kelempong | 2021 | ranjang |
| 123 | angkang | 756 | dade | 1389 | kelende | 2022 | ransang |
| 124 | angkaq | 757 | dagul | 1390 | kelenger | 2023 | ransek |
| 125 | angkat | 758 | dahar | 1391 | keleong | 2024 | rantek |
| 126 | angken | 759 | dahe | 1392 | kelepeng | 2025 | raok |

| | | | | | | | |
|-----|-------------------|-----|----------|------|----------|------|----------|
| 127 | angker | 760 | dait | 1393 | kelet | 2026 | raos |
| 128 | angkos | 761 | dakaq | 1394 | kelewang | 2027 | rapat |
| 129 | angku | 762 | dalang | 1395 | keliang | 2028 | rapek |
| 130 | angkus | 763 | dam | 1396 | keliang | 2029 | ngerapek |
| 131 | angkut | 764 | dames | 1397 | kelin | 2030 | rapet |
| 132 | angoh | 765 | dampeq | 1398 | keling | 2031 | rapus |
| 133 | angop | 766 | dampon | 1399 | kelipung | 2032 | raq |
| 134 | angse | 767 | dampuk | 1400 | kelir | 2033 | rarang |
| 135 | aniayaq | 768 | dandaq | 1401 | kelit | 2034 | raraq |
| 136 | aning | 769 | dangkem | 1402 | kelocok | 2035 | raras |
| 137 | anjani | 770 | daran | 1403 | kelor | 2036 | rarat |
| 138 | anjar | 771 | darag | 1404 | kelotih | 2037 | rarut |
| 139 | anjek | 772 | darat | 1405 | kelus | 2038 | rase |
| 140 | anjeng | 773 | dare | 1406 | kemaruk | 2039 | rase |
| 141 | nganjeng | 774 | dare | 1407 | kembah | 2040 | rasun |
| 142 | anjur | 775 | dedare | 1408 | kembak | 2041 | rat |
| 143 | ansen | 776 | dareq | 1409 | kembang | 2042 | rate |
| 144 | ansok | 777 | dares | 1410 | kembar | 2043 | rauh |
| 145 | ansor | 778 | das | 1411 | kemben | 2044 | rayuq |
| 146 | ansuh | 779 | dasan | 1412 | kembeq | 2045 | re |
| 147 | ansul | 780 | dase | 1413 | kemboq | 2046 | reban |
| 148 | ansur | 781 | dateng | 1414 | kembulan | 2047 | rebaq |
| 149 | ante | 782 | datoq | 1415 | kemelet | 2048 | regaq |
| 150 | anteh | 783 | datu | 1416 | kemeng | 2049 | regeng |
| 151 | antem | 784 | daun | 1417 | kemaq | 2050 | regu |
| 152 | antep | 785 | daun | 1418 | kemeras | 2051 | rekeng |
| 153 | antih | 786 | dawek | 1419 | kemeri | 2052 | rekok |
| 154 | anting- anting | 787 | daye | 1420 | kemos | 2053 | rebong |
| 155 | antiq | 788 | deang | 1421 | kemotoq | 2054 | rebot |
| 156 | anton | 789 | dede | 1422 | kempeng | 2055 | rempen |
| 157 | antos | 790 | dedek | 1423 | kempes | 2056 | rempet |
| 158 | antu | 791 | dedep | 1424 | kempuq | 2057 | reng |
| 159 | antuq | 792 | dee | 1425 | kenaq | 2058 | rengkos |
| 160 | anuq | 793 | bedee | 1426 | kencak | 2059 | repok |
| 161 | anyir | 794 | deket | 1427 | kenceng | 2060 | repus |
| 162 | aop | 795 | demak | 1428 | kendali | 2061 | rerek |
| 163 | aoq | 796 | demen | 1429 | kendur | 2062 | rereq |
| 164 | apah | 797 | dempe | 1430 | kene | 2063 | rerik |
| 165 | apak | 798 | dempek | 1431 | kenjan | 2064 | resit |
| 166 | apal | 799 | dendeq | 1432 | kenjeq | 2065 | ret |
| 167 | apaq | 800 | dendeq | 1433 | kentaq | 2066 | retak |
| 168 | ape | 801 | dengah | 1434 | kentok | 2067 | reteq |
| 169 | apeq | 802 | dengan | 1435 | kenyamen | 2068 | retoq |
| 170 | apik | 803 | denggong | 1436 | kenyeru | 2069 | ribut |

| | | | | | | | |
|-----|--------|-----|----------|------|----------|------|--------|
| 171 | apit | 804 | densrot | 1437 | kepak | 2070 | rimong |
| 172 | apuh | 805 | depak | 1438 | kepet | 2071 | ririh |
| 173 | apung | 806 | derep | 1439 | kepio | 2072 | ririk |
| 174 | apus | 807 | bederep | 1440 | kepot | 2073 | roah |
| 175 | arah | 808 | det | 1441 | keपुरi | 2074 | robek |
| 176 | arak | 809 | dilah | 1442 | keramak | 2075 | rokep |
| 177 | aran | 810 | dingang | 1443 | keraro | 2076 | romboq |
| 178 | arang | 811 | dingaq; | 1444 | kerah | 2077 | rong |
| 179 | araq | 812 | kedingaq | 1445 | keremon | 2078 | ronjoq |
| 180 | arat | 813 | direq | 1446 | kereng | 2079 | rontok |
| 181 | arem | 814 | dodoq | 1447 | kereng | 2080 | ropek |
| 182 | aren | 815 | dodot | 1448 | kerengeq | 2081 | roroh |
| 183 | areng | 816 | doe | 1449 | keroq | 2082 | rosep |
| 184 | areq | 817 | doe | 1450 | keret | 2083 | ruan |
| 185 | ares | 818 | dokar | 1451 | kering | 2084 | rubin |
| 186 | aro | 819 | donyok | 1452 | kerinsoq | 2085 | rue |
| 187 | aroh | 820 | dopok | 1453 | kerisaq | 2086 | rugl |
| 188 | arok | 821 | dore | 1454 | keriun | 2087 | rujuq |
| 189 | arong | 822 | doyan | 1455 | kerodok | 2088 | rukik |
| 190 | arot | 823 | dugaq | 1456 | keruan | 2089 | rumpak |
| 191 | arti | 824 | dukep | 1457 | kerujuq | 2090 | rundam |
| 192 | arus | 825 | dulang | 1458 | keruk | 2091 | runguq |
| 193 | arwah | 826 | dure | 1459 | kerut | 2092 | runtak |
| 194 | asah | 827 | durus | 1460 | kesanduk | 2093 | rupek |
| 195 | asal | 828 | dus | 1461 | keset | 2094 | rurung |
| 196 | asaq | 829 | eat | 1462 | kesur | 2095 | rurut |
| 197 | asar | 830 | edaq | 1463 | kesusu | 2096 | ruse |
| 198 | asbak | 831 | edeh | 1464 | kesuwu | 2097 | rusep |
| 199 | ase | 832 | ek | 1465 | ket | 2098 | sadu |
| 200 | ase | 833 | ekek | 1466 | ketaon | 2099 | sae |
| 201 | ngase | 834 | eket | 1467 | ketemuq | 2100 | saget |
| 202 | asem | 835 | elah | 1468 | keterung | 2101 | sah |
| 203 | aseq | 836 | elang | 1469 | kewaq | 2102 | sah |
| 204 | asin | 837 | eleh | 1470 | kiblas | 2103 | sait |
| 205 | astage | 838 | elek | 1471 | kidung | 2104 | kesait |
| 206 | asuj | 839 | elen | 1472 | kikit | 2105 | sak |
| 207 | atas | 840 | eloq | 1473 | kileng | 2106 | sakit |
| 208 | atawe | 841 | seloq | 1474 | kintur | 2107 | salam |
| 209 | ate | 842 | elor | 1475 | kisut | 2108 | salaq |
| 210 | atep | 843 | embas | 1476 | kocet | 2109 | sale |
| 211 | ater | 844 | embek | 1477 | kocok | 2110 | saleh |
| 212 | atom | 845 | emben | 1478 | kocor | 2111 | salin |
| 213 | atong | 846 | embeng | 1479 | kodeq | 2112 | salo |
| 214 | atuh | 847 | emboh | 1480 | koeh | 2113 | saluk |

| | | | | | | | |
|-----|-------------|-----|-------------|------|---------|------|----------|
| 215 | atur | 848 | embot | 1481 | kojoh | 2114 | sambang |
| 216 | aur | 849 | embuk | 1482 | kojong | 2115 | samber |
| 217 | aus | 850 | ngembuk | 1483 | kojor | 2116 | sambi |
| 218 | aut | 851 | embun | 1484 | kokoh | 2117 | sambung |
| 219 | awak | 852 | embung | 1485 | koloq | 2118 | sambut |
| 220 | awak | 853 | embur | 1486 | kombet | 2119 | sampah |
| 221 | ngawak | 854 | embur-embur | 1487 | kong | 2120 | sampak |
| 222 | awal | 855 | embus | 1488 | kongon | 2121 | sampan |
| 223 | awam | 856 | emoq | 1489 | kongong | 2122 | sampat |
| 224 | awan | 857 | empah | 1490 | konteq | 2123 | sampeq |
| 225 | awang-awang | 858 | empang | 1491 | konyaq | 2124 | sampun |
| 226 | awas | 859 | empaq | 1492 | kopok | 2125 | sampunan |
| 227 | awek | 860 | empas | 1493 | koras | 2126 | sandaq |
| 228 | awet | 861 | emper | 1494 | kotok | 2127 | sandel |
| 229 | awik | 862 | empes | 1495 | kotong | 2128 | sang |
| 230 | awik-awik | 863 | empet | 1496 | kotor | 2129 | sange |
| 231 | awin | 864 | empiq | 1497 | koyan | 2130 | sanget |
| 232 | awon | 865 | empoh | 1498 | koyok | 2131 | sanggup |
| 233 | awon-awon | 866 | empok | 1499 | koyos | 2132 | sangke |
| 234 | awor | 867 | empoq | 1500 | kuali | 2133 | sangkep |
| 235 | awuk-awuk | 868 | empuk | 1501 | kuat | 2134 | sangkep |
| 236 | awun-awun | 869 | endah | 1502 | kudung | 2135 | sangklap |
| 237 | ayah | 870 | endan | 1503 | kueli | 2136 | sangkor |
| 238 | ngayah | 871 | endaus | 1504 | kuih | 2137 | sangkur |
| 239 | ayaq | 872 | ende | 1505 | kuih | 2138 | sango |
| 240 | ayat | 873 | ende | 1506 | bekuih | 2139 | sante |
| 241 | babah | 874 | endeng | 1507 | kulak | 2140 | santer |
| 242 | babak | 875 | endeq | 1508 | kuliq | 2141 | saok |
| 243 | babar | 876 | endes | 1509 | kurek | 2142 | sap |
| 244 | babas | 877 | endet | 1510 | kuren | 2143 | sapak |
| 245 | babat | 878 | endom | 1511 | kuri | 2144 | sapaq |
| 246 | bacem | 879 | endon | 1512 | kuris | 2145 | saprek |
| 247 | bacot | 880 | ngendon | 1513 | kusem | 2146 | saprih |
| 248 | badaq | 881 | endot | 1514 | kut | 2147 | sapu |
| 249 | badeq | 882 | enduk | 1515 | kutek | 2148 | saq |
| 250 | badiq | 883 | endut | 1516 | kuwawe | 2149 | sarak |
| 251 | badon | 884 | engat | 1517 | kuwur | 2150 | goah |
| 252 | baduk | 885 | enges | 1518 | labir | 2151 | sarak |
| 253 | badung | 886 | enget | 1519 | labuh | 2152 | sarung |
| 254 | badut | 887 | enggaq | 1520 | labur | 2153 | sato |
| 255 | bae | 888 | enggep | 1521 | laci | 2154 | satuq |
| 256 | baeh | 889 | engger | 1522 | lacur | 2155 | saur |
| 257 | baes | 890 | enggih | 1523 | ladek | 2156 | saut |
| 258 | bagi | 891 | engkah | 1524 | lader | 2157 | sawe |

| | | | | | | | |
|-----|---------|-----|-----------|------|------------|------|-------------|
| 259 | bagiq | 892 | engkah | 1525 | laeq | 2158 | seang |
| 260 | bagu | 893 | engkar | 1526 | laet | 2159 | sebang |
| 261 | bagus | 894 | engoh | 1527 | lagan | 2160 | sebit |
| 262 | bah | 895 | engos | 1528 | lagar | 2161 | sebuq |
| 263 | bahle | 896 | enjeq | 1529 | lage | 2162 | sebur |
| 264 | bai | 897 | entak | 1530 | lain | 2163 | sebut |
| 265 | bais | 898 | entan | 1531 | laiq | 2164 | sedaq |
| 266 | bait | 899 | enteh | 1532 | lakaq | 2165 | sedaq |
| 267 | bajang | 900 | entek | 1533 | latakaq | 2166 | sede |
| 268 | baje | 901 | entem | 1534 | lakar | 2167 | sede |
| 269 | bajo | 902 | enter | 1535 | lakong | 2168 | sedeng |
| 270 | bajur | 903 | entir | 1536 | lakoq | 2169 | sediah |
| 271 | bak | 904 | entong | 1537 | lalah | 2170 | sedut |
| 272 | bakal | 905 | entuk | 1538 | pelalah | 2171 | seger |
| 273 | bakaq | 906 | entur | 1539 | lalang | 2172 | segerah |
| 274 | bakat | 907 | epe | 1540 | lalo | 2173 | seher |
| 275 | bakeq | 908 | epe | 1541 | laloq | 2174 | sekat |
| 276 | baker | 909 | epot | 1542 | lamak | 2175 | sekediq |
| 277 | bakti | 910 | eraq | 1543 | lamang | 2176 | seken |
| 278 | bal | 911 | eres | 1544 | ngelamang | 2177 | sekep |
| 279 | balar | 912 | esot | 1545 | lamar | 2178 | sekeq |
| 280 | balas | 913 | gabah | 1546 | lamas | 2179 | sekoq |
| 281 | bale | 914 | gabak | 1547 | lambah | 2180 | selabek |
| 282 | balek | 915 | gabeng | 1548 | lambang | 2181 | selah |
| 283 | balet | 916 | gabung | 1549 | lames | 2182 | selak |
| 284 | balit | 917 | gacek | 1550 | lampak | 2183 | selamian |
| 285 | kembali | 918 | gading | 1551 | lampaq | 2184 | selane |
| 286 | balok | 919 | gagah | 1552 | lamper | 2185 | selao |
| 287 | balon | 920 | gagak | 1553 | lampin | 2186 | selapuq |
| 288 | baloq | 921 | gagal | 1554 | lancar | 2187 | selaq |
| 289 | balung | 922 | gaji | 1555 | belancaran | 2188 | seledok |
| 290 | baluq | 923 | galah | 1556 | langah | 2189 | seleme |
| 291 | bamban | 924 | galak | 1557 | langan | 2190 | seleng |
| 292 | ban | 925 | galang | 1558 | langger | 2191 | selet |
| 293 | bande | 926 | gale | 1559 | langgiq | 2192 | selewoq |
| 294 | bandel | 927 | gale-gale | 1560 | lanjak | 2193 | seloar |
| 295 | bandeng | 928 | galeng | 1561 | lanjaq | 2194 | seluk |
| 296 | banding | 929 | gaet | 1562 | lanset | 2195 | selung |
| 297 | bandung | 930 | galip | 1563 | lansot | 2196 | selung-lung |
| 298 | bang | 931 | galir | 1564 | lantih | 2197 | sembah |
| 299 | bangaq | 932 | galuh | 1565 | lantong | 2198 | sembar |
| 300 | banger | 933 | galur | 1566 | lantong | 2199 | sembih |
| 301 | banget | 934 | gamaq | 1567 | lantur | 2200 | semel |
| 302 | bangge | 935 | gambar | 1568 | lapah | 2201 | semeton |

| | | | | | | | |
|-----|-----------|-----|-----------|------|------------|------|------------|
| 303 | bangruq | 936 | gambik | 1569 | lapang | 2202 | sempait |
| 304 | bangkal | 937 | gampung | 1570 | lapas | 2203 | semper |
| 305 | bangkang | 938 | gancuh | 1571 | lapis | 2204 | semperak |
| 306 | bangke | 939 | gandong | 1572 | lapor | 2205 | semprot |
| 307 | bangkel | 940 | ganggas | 1573 | latah | 2206 | semu |
| 308 | bangkes | 941 | ganggu | 1574 | laun | 2207 | seneng |
| 309 | bangket | 942 | ganjaq | 1575 | lawan | 2208 | sengam |
| 310 | bangkol | 943 | gantang | 1576 | lawang | 2209 | senggeq |
| 311 | bangkoq | 944 | gantoq | 1577 | lawas | 2210 | senggitan |
| 312 | bangkrut | 945 | gantung | 1578 | lawoq | 2211 | sengit |
| 313 | bangkung | 946 | gantus | 1579 | leang | 2212 | sengker |
| 314 | bangsat | 947 | gaos | 1580 | leancang | 2213 | sengker |
| 315 | bangse | 948 | gapit | 1581 | keleancang | 2214 | senjate |
| 316 | bangun | 949 | gare-gare | 1582 | lecok | 2215 | senoq |
| 317 | bani | 950 | garis | 1583 | lege | 2216 | sentulak |
| 318 | banjah | 951 | garo | 1584 | leger | 2217 | besentulak |
| 319 | banjar | 952 | garok | 1585 | lekak | 2218 | senuq |
| 320 | bansu | 953 | garus | 1586 | lekan | 2219 | sep |
| 321 | bantal | 954 | garut | 1587 | lekaq | 2220 | sepaq |
| 322 | bantar | 955 | gasap | 1588 | lekaq | 2221 | sepet |
| 323 | bantat | 956 | gase | 1589 | lekes | 2222 | sepi |
| 324 | bantel | 957 | gati | 1590 | leket | 2223 | sepoq |
| 325 | banteng | 958 | gawat | 1591 | lekoq | 2224 | sepor |
| 326 | banting | 959 | gawe | 1592 | lelah | 2225 | sepuq |
| 327 | bantot | 960 | gaweq | 1593 | lelang | 2226 | serah |
| 328 | bantu | 961 | gaweq | 1594 | leleh | 2227 | seran |
| 329 | banyol | 962 | gawik | 1595 | leleq | 2228 | seraq |
| 330 | bao | 963 | gawos | 1596 | leles | 2229 | sere |
| 331 | baoq | 964 | gayas | 1597 | lelet | 2230 | sere |
| 332 | baos | 965 | gaye | 1598 | lemak | 2231 | serek |
| 333 | bapaq | 966 | gebah | 1599 | leman | 2232 | sereng |
| 334 | bape | 967 | gebar | 1600 | lemaq | 2233 | serep |
| 335 | barak | 968 | gebut | 1601 | lembah | 2234 | serepek |
| 336 | barang | 969 | gebuk | 1602 | lemboq | 2235 | sereq |
| 337 | baraq | 970 | gecek | 1603 | lempek | 2236 | seret |
| 338 | barat | 971 | gecik | 1604 | lempen | 2237 | seri |
| 339 | bare | 972 | gecok | 1605 | lempot | 2238 | serikaye |
| 340 | bare-bare | 973 | gedek | 1606 | lemu | 2239 | serogoq |
| 341 | bareh | 974 | gedeng | 1607 | lencos | 2240 | set |
| 342 | bareng | 975 | gedok | 1608 | lendang | 2241 | setan |
| 343 | baret | 976 | gedoq | 1609 | lendek | 2242 | setir |
| 344 | bari | 977 | begedoq | 1610 | lendong | 2243 | setor |
| 345 | baris | 978 | gedor | 1611 | lengeq | 2244 | sewe |
| 346 | barungan | 979 | gejuh | 1612 | lengker | 2245 | siak |

| | | | | | | | |
|-----|---------|------|--------------|------|----------|------|---------------|
| 347 | baruq | 980 | begejuh | 1613 | lenjang | 2246 | siar |
| 348 | barus | 981 | gelampar | 1614 | lenting | 2247 | sibuk |
| 349 | basaq | 982 | begelamparan | 1615 | lenting | 2248 | side |
| 350 | basar | 983 | gelanyat | 1616 | lentos | 2249 | sidem |
| 351 | base | 984 | begelanyat | 1617 | leor | 2250 | sigon |
| 352 | basong | 985 | gelaq | 1618 | leot | 2251 | sik |
| 353 | bat | 986 | gelat | 1619 | lepah | 2252 | sikep |
| 354 | batah | 987 | geles | 1620 | lepong | 2253 | sikut |
| 355 | batal | 988 | gelompong | 1621 | lepaq | 2254 | silaq |
| 356 | batang | 989 | gelung | 1622 | lepas | 2255 | sile |
| 357 | batar | 990 | gelut | 1623 | lepek | 2256 | sili |
| 358 | batbat | 991 | gemes | 1624 | lepoh | 2257 | silung-silung |
| 359 | bateq | 992 | gendang | 1625 | longkop | 2258 | silur |
| 360 | bates | 993 | gendaq | 1626 | lonjong | 2259 | simbut |
| 361 | bati | 994 | gending | 1627 | lonjor | 2260 | simo |
| 362 | batuk | 995 | gendong | 1628 | lontah | 2261 | sindang |
| 363 | batur | 996 | gendut | 1629 | lontah | 2262 | singgaq |
| 364 | bau | 997 | genem | 1630 | lontas | 2263 | sintek |
| 365 | baur | 998 | genep | 1631 | lose | 2264 | siq |
| 366 | baut | 999 | genit | 1632 | loset | 2265 | sirep |
| 367 | bawaq | 1000 | genjah | 1633 | lotoh | 2266 | siru |
| 368 | bawi | 1001 | gentiq | 1634 | lowok | 2267 | siruk |
| 369 | bawo | 1002 | gerame | 1635 | loyo | 2268 | sise |
| 370 | bawot | 1003 | gerak | 1636 | luap | 2269 | sisik |
| 371 | bayem | 1004 | gereng | 1637 | luar | 2270 | siteng |
| 372 | bayoq | 1005 | gereot | 1638 | lueq | 2271 | siuh |
| 373 | bayu | 1006 | gerepek | 1639 | luken | 2272 | sodoa |
| 374 | be | 1007 | begerepek | 1640 | lukeq | 2273 | soek |
| 375 | be | 1008 | gerepet | 1641 | lukes | 2274 | solah |
| 376 | beang | 1009 | begerepet | 1642 | lumbar | 2275 | solong |
| 377 | beaq | 1010 | geres | 1643 | lunas | 2276 | sombon |
| 378 | bebeaq | 1011 | geret | 1644 | luncer | 2277 | sombong |
| 379 | bebadas | 1012 | gergaji | 1645 | lungkup | 2278 | sompel |
| 380 | bebai | 1013 | gerik | 1646 | lunsuran | 2279 | songket |
| 381 | bebante | 1014 | gering | 1647 | lupaq | 2280 | songkok |
| 382 | bebas | 1015 | geriq | 1648 | luten | 2281 | songkoq |
| 383 | bebek | 1016 | gero | 1649 | maap | 2282 | songo |
| 384 | bebel | 1017 | gerobak | 1650 | madeq | 2283 | sopoq |
| 385 | beber | 1018 | geroh | 1651 | mah | 2284 | sorah |
| 386 | bebet | 1019 | gerome | 1652 | mahir | 2285 | sorok |
| 387 | bebetok | 1020 | geropak | 1653 | maiq | 2286 | sorong |
| 388 | becang | 1021 | begeropak | 1654 | maksut | 2287 | sripe |
| 389 | becaq | 1022 | gerup | 1655 | mal | 2288 | suap |
| 390 | becat | 1023 | gerupuk | 1656 | malah | 2289 | subruk |

| | | | | | | | |
|-----|-----------|------|-----------|------|---------|------|----------|
| 391 | beciq | 1024 | gerus | 1657 | malang | 2290 | subuh |
| 392 | bedah | 1025 | gerus | 1658 | male | 2291 | sugi |
| 393 | bedak | 1026 | gres | 1659 | malem | 2292 | sugih |
| 394 | bede | 1027 | gesik | 1660 | males | 2293 | sumbang |
| 395 | bedek | 1028 | gesit | 1661 | malik | 2294 | sumbar |
| 396 | bedel | 1029 | getas | 1662 | malit | 2295 | sumpel |
| 397 | bedeng | 1030 | geteng | 1663 | mambis | 2296 | sungkan |
| 398 | bedil | 1031 | geti | 1664 | mamet | 2297 | sungkep |
| 399 | bedu | 1032 | getok | 1665 | mamiq | 2298 | sungkiq |
| 400 | beduk | 1033 | gewart | 1666 | mampu | 2299 | supuq |
| 401 | bedul | 1034 | gigi | 1667 | man | 2300 | surah |
| 402 | bedus | 1035 | gile | 1668 | mance | 2301 | surak |
| 403 | begal | 1036 | giling | 1669 | mancer | 2302 | susur |
| 404 | begang | 1037 | gilir | 1670 | mancet | 2303 | susut |
| 405 | beqaq | 1038 | giong | 1671 | mancung | 2304 | sut |
| 406 | bege | 1039 | girang | 1672 | mancur | 2305 | suter |
| 407 | bezik | 1040 | giras | 1673 | mandi | 2306 | tabah |
| 408 | bejat | 1041 | begiras | 1674 | mandiq | 2307 | tabeq |
| 409 | bejek | 1042 | gitaq | 1675 | mangan | 2308 | tadah |
| 410 | bekaq | 1043 | goah | 1676 | mangkin | 2309 | tadong |
| 411 | bekas | 1044 | goap | 1677 | manik | 2310 | tahen |
| 412 | bekeh | 1045 | goar | 1678 | manse | 2311 | tai |
| 413 | bekel | 1046 | gocek | 1679 | mantuk | 2312 | tajem |
| 414 | bekeq | 1047 | godek | 1680 | manuh | 2313 | tajuk |
| 415 | bekeq | 1048 | goloh | 1681 | maqum | 2314 | takaq |
| 416 | beket | 1049 | gompar | 1682 | maraq | 2315 | taker |
| 417 | bekot | 1050 | goret | 1683 | mare | 2316 | takut |
| 418 | belabur | 1051 | goroh | 1684 | mareq | 2317 | talang |
| 419 | belah | 1052 | gorok | 1685 | mase | 2318 | talet |
| 420 | belang | 1053 | gorong | 1686 | mate | 2319 | talo |
| 421 | belang | 1054 | got | 1687 | mate | 2320 | taloq |
| 422 | belangsir | 1055 | gubah | 1688 | mating | 2321 | tamaq |
| 423 | belantuk | 1056 | gubuk | 1689 | mauq | 2322 | tamat |
| 424 | belaq | 1057 | gulah | 1690 | mayang | 2323 | tambah |
| 425 | belar | 1058 | gule | 1691 | mayung | 2324 | tambang |
| 426 | belas | 1059 | gulem | 1692 | mayus | 2325 | tampe |
| 427 | belat | 1060 | gulung | 1693 | me | 2326 | tampeng |
| 428 | belate | 1061 | gumane | 1694 | meco | 2327 | tamput |
| 429 | belatung | 1062 | gumbleng | 1695 | meje | 2328 | tanaq |
| 430 | bele | 1063 | gumi | 1696 | meke | 2329 | tandak |
| 431 | belek | 1064 | gumi | 1697 | mel | 2330 | betandak |
| 432 | belek | 1065 | paer | 1698 | melak | 2331 | tandang |
| 433 | belenggu | 1066 | gune | 1699 | melaq | 2332 | tande |
| 434 | beleq | 1067 | gune-gune | 1700 | melaq | 2333 | tandoq |

| | | | | | | | |
|-----|-----------|------|---------|------|------------|------|----------|
| 435 | beler | 1068 | guntur | 1701 | mele | 2334 | tanggeq |
| 436 | belet | 1069 | gupuh | 1702 | melet | 2335 | tanggor |
| 437 | belian | 1070 | guru | 1703 | meletek | 2336 | tangis |
| 438 | belide | 1071 | hadir | 1704 | melong | 2337 | tangkis |
| 439 | belikat | 1072 | hadis | 1705 | melut | 2338 | tangkong |
| 440 | belinggi | 1073 | haji | 1706 | memes | 2339 | tango |
| 441 | belinjo | 1074 | hak | 1707 | men | 2340 | tanjaq |
| 442 | belis | 1075 | hakekat | 1708 | men | 2341 | tanjek |
| 443 | belitur | 1076 | halal | 1709 | menah | 2342 | tao |
| 444 | beliung | 1077 | hampir | 1710 | mendot | 2343 | tapak |
| 445 | belo | 1078 | haram | 1711 | meneh | 2344 | tape |
| 446 | beloh | 1079 | harep | 1712 | menge | 2345 | tarik |
| 447 | belok | 1080 | harte | 1713 | menggah | 2346 | taring |
| 448 | belok | 1081 | harus | 1714 | mengkang | 2347 | tatu |
| 449 | belondan | 1082 | hasil | 1715 | kemengkang | 2348 | tau |
| 450 | belong | 1083 | hebat | 1716 | meni | 2349 | te |
| 451 | belot | 1084 | heler | 1717 | meno | 2350 | te |
| 452 | belot | 1085 | het | 1718 | mensang | 2351 | tebaq |
| 453 | bembeq | 1086 | hindu | 1719 | mensang | 2352 | tebok |
| 454 | benah | 1087 | hut | 1720 | menu | 2353 | tedem |
| 455 | bendang | 1088 | ibadah | 1721 | menu | 2354 | tedok |
| 456 | bendem | 1089 | ibarat | 1722 | menyah | 2355 | tedong |
| 457 | bender | 1090 | ican | 1723 | merang | 2356 | tedoq |
| 458 | bendung | 1091 | idap | 1724 | merariq | 2357 | tegel |
| 459 | beneng | 1092 | ider | 1725 | merebek | 2358 | teguh |
| 460 | benes | 1093 | idiq | 1726 | mereng | 2359 | teguq |
| 461 | beng | 1094 | igel | 1727 | meres | 2360 | tegur |
| 462 | bengan | 1095 | ihlas | 1728 | meriaq | 2361 | tehel |
| 463 | beng-beng | 1096 | ijeng | 1729 | meriku | 2362 | tekak |
| 464 | bengel | 1097 | ikel | 1730 | meriq | 2363 | teken |
| 465 | benger | 1098 | iket | 1731 | mero | 2364 | teker |
| 466 | bengis | 1099 | ilip | 1732 | mertaq | 2365 | tekes |
| 467 | bengkes | 1100 | ilmu | 1733 | merubuk | 2366 | tekiq |
| 468 | bengkis | 1101 | ilon | 1734 | met | 2367 | tekor |
| 469 | bengkok | 1102 | imam | 1735 | metu | 2368 | tekus |
| 470 | bengkudu | 1103 | iman | 1736 | milu | 2369 | telabah |
| 471 | bengkung | 1104 | ime | 1737 | mimis | 2370 | telang |
| 472 | benguh | 1105 | impan | 1738 | mimit | 2371 | telat |
| 473 | benguq | 1106 | impung | 1739 | mindah | 2372 | tele |
| 474 | benjor | 1107 | inaq | 1740 | mintol | 2373 | tele |
| 475 | benjut | 1108 | incer | 1741 | modok | 2374 | telep |
| 476 | bentek | 1109 | indang | 1742 | moge | 2375 | telih |
| 477 | bentol | 1110 | ine | 1743 | mokoh | 2376 | telik |
| 478 | bentot | 1111 | inges | 1744 | molah | 2377 | telinjaq |

| | | | | | | | |
|-----|-----------|------|-----------|------|---------|------|-----------|
| 479 | berang | 1112 | inget | 1745 | momot | 2378 | telon |
| 480 | berani | 1113 | inggas | 1746 | monce | 2379 | telor |
| 481 | beras | 1114 | inggur | 1747 | mones | 2380 | telosor |
| 482 | berat | 1115 | iniq | 1748 | mongken | 2381 | tembok |
| 483 | berayan | 1116 | insah | 1749 | montong | 2382 | temerene |
| 484 | beraye | 1117 | insap | 1750 | monyan | 2383 | temerubuk |
| 485 | berek | 1118 | inte | 1751 | morot | 2384 | temin |
| 486 | berem | 1119 | intek | 1752 | mose | 2385 | tempang |
| 487 | berembe | 1120 | intem | 1753 | mosot | 2386 | tampe |
| 488 | beren | 1121 | ipuh | 1754 | mot | 2387 | tempengel |
| 489 | berengkes | 1122 | irap | 1755 | mudaq | 2388 | tenang |
| 490 | berengos | 1123 | iring | 1756 | mue | 2389 | tenaq |
| 491 | beres | 1124 | iris | 1757 | mujuk | 2390 | tendang |
| 492 | beres | 1125 | irit | 1758 | mujur | 2391 | tender |
| 493 | beberes | 1126 | iroq | 1759 | mukur | 2392 | tendes |
| 494 | bergu | 1127 | irung | 1760 | mule | 2393 | tendos |
| 495 | berhale | 1128 | irup | 1761 | mulus | 2394 | tengari |
| 496 | beringin | 1129 | irus | 1762 | mulut | 2395 | tengeru |
| 497 | beriq | 1130 | isiq | 1763 | muri | 2396 | tenget |
| 498 | berire | 1131 | isoq | 1764 | muwuk | 2397 | tenggale |
| 499 | beriuq | 1132 | itam | 1765 | nae | 2398 | tenggong |
| 500 | berkat | 1133 | iter | 1766 | nani | 2399 | tengin |
| 501 | beroh | 1134 | itiq | 1767 | nare | 2400 | tenjot |
| 502 | berong | 1135 | itiq-itiq | 1768 | nasiq | 2401 | tenun |
| 503 | berongge | 1136 | itung | 1769 | nat | 2402 | tenung |
| 504 | beroq | 1137 | iur | 1770 | ne | 2403 | tepek |
| 505 | beberoq | 1138 | iwong | 1771 | nejis | 2404 | tepeng |
| 506 | berorong | 1139 | iwoq | 1772 | nekat | 2405 | tepes |
| 507 | bersih | 1140 | jabin | 1773 | neneq | 2406 | tepes |
| 508 | berubusan | 1141 | jabrik | 1774 | nenten | 2407 | tepok |
| 509 | beruqaq | 1142 | jabung | 1775 | nepar | 2408 | tepong |
| 510 | beruge | 1143 | jabut | 1776 | ngalaq | 2409 | terimaq |
| 511 | berumbun | 1144 | jadah | 1777 | ngandot | 2410 | teriq |
| 512 | besbes | 1145 | jadik | 1778 | ngangaq | 2411 | terompe |
| 513 | besoq | 1146 | jagaq | 1779 | ngantok | 2412 | terune |
| 514 | besuh | 1147 | jage | 1780 | ngege | 2413 | terutuk |
| 515 | betah | 1148 | jage | 1781 | ngemeq | 2414 | tetu |
| 516 | betaq | 1149 | jago | 1782 | ngemeq | 2415 | tian |
| 517 | betaran | 1150 | jagur | 1783 | ngengeq | 2416 | tiang |
| 518 | betek | 1151 | jaket | 1784 | ngerep | 2417 | ticaq |
| 519 | betel | 1152 | jale | 1785 | ngerok | 2418 | tidem |
| 520 | betelah | 1153 | jaler | 1786 | ni | 2419 | tikah |
| 521 | betoq | 1154 | jam | 1787 | nie | 2420 | tiken |
| 522 | betuk | 1155 | jaman | 1788 | nike | 2421 | tikus |

| | | | | | | | |
|-----|---------|------|-------------|------|-----------|------|----------|
| 523 | bewe | 1156 | jamaq | 1789 | niki | 2422 | tilah |
| 524 | bi | 1157 | james | 1790 | nine | 2423 | tilas |
| 525 | biah | 1158 | kejames | 1791 | niniq | 2424 | timaq |
| 526 | biak | 1159 | jangan | 1792 | niste | 2425 | timbal |
| 527 | bian | 1160 | jangke | 1793 | no | 2426 | timbaq |
| 528 | biar | 1161 | jangkit | 1794 | noaq | 2427 | timpoh |
| 529 | kebiar | 1162 | bejangkitan | 1795 | nu | 2428 | tindes |
| 530 | bibir | 1163 | jangkong | 1796 | nur | 2429 | tindih |
| 531 | bideng | 1164 | jango | 1797 | nurge | 2430 | tindoq |
| 532 | bih | 1165 | jangoq | 1798 | nyalah | 2431 | tinggang |
| 533 | bijang | 1166 | jangur | 1799 | nyale | 2432 | tingket |
| 534 | bije | 1167 | janji | 1800 | nyalung | 2433 | tinjal |
| 535 | biji | 1168 | jap | 1801 | nyangget | 2434 | tinjot |
| 536 | bikan | 1169 | jarah | 1802 | nyanyi | 2435 | tipah |
| 537 | bikas | 1170 | jarang | 1803 | nyariq | 2436 | tipaq |
| 538 | bikuq | 1171 | jaring | 1804 | nyat | 2437 | tipes |
| 539 | bilal | 1172 | jarup | 1805 | nyate | 2438 | tipis |
| 540 | bilang | 1173 | jati | 1806 | nyatoq | 2439 | tipu |
| 541 | bilaq | 1174 | jauq | 1807 | nyawang | 2440 | tirem |
| 542 | bilin | 1175 | jawang | 1808 | nyawung | 2441 | tirok |
| 543 | bilin | 1176 | jawab | 1809 | nyelili | 2442 | tiris |
| 544 | bebilin | 1177 | jaye | 1810 | nyenye | 2443 | titoq |
| 545 | bilok | 1178 | jebak | 1811 | nyerek | 2444 | tiup |
| 546 | biluk | 1179 | jebel | 1812 | nyeri | 2445 | tiwang |
| 547 | bineq | 1180 | jebir | 1813 | nyering | 2446 | tiwas |
| 548 | binger | 1181 | jebuk | 1814 | kenyering | 2447 | tiwoq |
| 549 | bingkal | 1182 | jeding | 1815 | nyet | 2448 | to |
| 550 | bingkir | 1183 | jegol | 1816 | nyieur | 2449 | toaq |
| 551 | bini | 1184 | jejah | 1817 | nyoer | 2450 | tua |
| 552 | biniq | 1185 | jeje | 1818 | nyonyong | 2451 | toeq |
| 553 | binjat | 1186 | jejeh | 1819 | nyunsaq | 2452 | toes |
| 554 | binteq | 1187 | jejet | 1820 | oap | 2453 | tokek |
| 555 | bintit | 1188 | jejuru | 1821 | oas | 2454 | tokol |
| 556 | bintur | 1189 | jelamar | 1822 | odaq | 2455 | tolang |
| 557 | biole | 1190 | jelantah | 1823 | odop | 2456 | toleq |
| 558 | birak | 1191 | jelap | 1824 | odor | 2457 | tongkem |
| 559 | biras | 1192 | jelas | 1825 | oet | 2458 | tonjok |
| 560 | bireng | 1193 | jelateng | 1826 | okep | 2459 | topat |
| 561 | biris | 1194 | jelek | 1827 | oleq | 2460 | tuak |
| 562 | bis | 1195 | jeleng | 1828 | oles | 2461 | manis |
| 563 | bisoq | 1196 | jeler | 1829 | ome | 2462 | tuak |
| 564 | bitek | 1197 | jelet | 1830 | omeh | 2463 | tuan |
| 565 | bitur | 1198 | jelik | 1831 | ompeh | 2464 | tuaq |
| 566 | biur | 1199 | jeloh | 1832 | ompet | 2465 | tugas |

| | | | | | | | |
|-----|----------|------|-----------|------|----------|------|----------|
| 567 | bius | 1200 | jelok | 1833 | ones | 2466 | tujaq |
| 568 | biwih | 1201 | jejelck | 1834 | ongkos | 2467 | tukah |
| 569 | bo | 1202 | jeluang | 1835 | ongoq | 2468 | tukaq |
| 570 | bobok | 1203 | jelut | 1836 | onjer | 2469 | tulak |
| 571 | boboq | 1204 | jemaq | 1837 | onos | 2470 | tumbas |
| 572 | bodak | 1205 | jemet | 1838 | onseng | 2471 | tumpes |
| 573 | bodo | 1206 | jemet | 1839 | onyaq | 2472 | tunah |
| 574 | bokbok | 1207 | jempang | 1840 | onyeh | 2473 | tunas |
| 575 | boke | 1208 | jemperang | 1841 | ope | 2474 | tunduk |
| 576 | bokon | 1209 | jempol | 1842 | opek | 2475 | tundung |
| 577 | bokoq | 1210 | jempulung | 1843 | opong | 2476 | tunggang |
| 578 | bokos | 1211 | jeneng | 1844 | opor | 2477 | tunggu |
| 579 | bolang | 1212 | jengah | 1845 | orah | 2478 | tungkem |
| 580 | bolaq | 1213 | jengeq | 1846 | oras | 2479 | tunjuk |
| 581 | bolet | 1214 | jenget | 1847 | ore | 2480 | tures |
| 582 | bolong | 1215 | jengget | 1848 | orong | 2481 | turis |
| 583 | bombong | 1216 | jenggit | 1849 | oros | 2482 | turun |
| 584 | bon | 1217 | jengking | 1850 | osok | 2483 | tutuq |
| 585 | bonder | 1218 | jepeng | 1851 | osok | 2484 | tuwuk |
| 586 | bong | 1219 | jepit | 1852 | otak | 2485 | uah |
| 587 | bonges | 1220 | jeput | 1853 | pacu | 2486 | uban |
| 588 | bongkang | 1221 | jeraq | 1854 | pade | 2487 | uber |
| 589 | bongkel | 1222 | jere | 1855 | paer | 2488 | uiq |
| 590 | bongkem | 1223 | jerezit | 1856 | pagah | 2489 | ujat |
| 591 | bongkor | 1224 | jeret | 1857 | pager | 2490 | uken |
| 592 | bongkos | 1225 | jerih | 1858 | paice | 2491 | ukep |
| 593 | bongkot | 1226 | jering | 1859 | pait | 2492 | ulat |
| 594 | bontet | 1227 | jero | 1860 | pajar | 2493 | ulem |
| 595 | bontet | 1228 | jerong | 1861 | pajek | 2494 | uleq |
| 596 | bonto | 1229 | jeruk | 1862 | pakoq | 2495 | ules |
| 597 | bebonto | 1230 | jerupi | 1863 | paku | 2496 | uli |
| 598 | botong | 1231 | jeruwuk | 1864 | palang | 2497 | ulu |
| 599 | bonyeq | 1232 | bejeruwuk | 1865 | paleq | 2498 | umah |
| 600 | bor | 1233 | jetit | 1866 | paling | 2499 | umaq |
| 601 | borang | 1234 | jeweh | 1867 | panas | 2500 | umat |
| 602 | boreh | 1235 | jewel | 1868 | panci | 2501 | umpet |
| 603 | borek | 1236 | jeye | 1869 | pangan | 2502 | unggat |
| 604 | borong | 1237 | jim | 1870 | pangeran | 2503 | uni |
| 605 | boros | 1238 | jimat | 1871 | panggal | 2504 | uni |
| 606 | borot | 1239 | jin | 1872 | panjak | 2505 | unjer |
| 607 | bos | 1240 | jinjit | 1873 | panjar | 2506 | unseng |
| 608 | bosok | 1241 | jiwang | 1874 | pansek | 2507 | untung |
| 609 | bosokan | 1242 | jiwe | 1875 | pantek | 2508 | upaq |
| 610 | bosor | 1243 | jiwel | 1876 | pantok | 2509 | urak |

| | | | | | | | |
|-----|---------|------|----------|------|-----------|------|---------|
| 611 | botak | 1244 | jiwit | 1877 | papah | 2510 | uras |
| 612 | botaq | 1245 | joang | 1878 | papaq | 2511 | ure |
| 613 | boteq | 1246 | joat | 1879 | parampuan | 2512 | usap |
| 614 | botes | 1247 | jogang | 1880 | paran | 2513 | usut |
| 615 | botoh | 1248 | joget | 1881 | parap | 2514 | utaq |
| 616 | bebotoh | 1249 | jojoq | 1882 | paras | 2515 | uwus |
| 617 | botor | 1250 | jolak | 1883 | parek | 2516 | uyut |
| 618 | bowos | 1251 | jole | 1884 | pariq | 2517 | uzur |
| 619 | boyaq | 1252 | jomet | 1885 | paro | 2518 | waje |
| 620 | boye | 1253 | jongger | 1886 | pas | 2519 | wajik |
| 621 | boyong | 1254 | jongjong | 1887 | pasah | 2520 | wajip |
| 622 | boyot | 1255 | jongkang | 1888 | pasengan | 2521 | waktu |
| 623 | buas | 1256 | jongkong | 1889 | pasti | 2522 | wanen |
| 624 | buat | 1257 | jongkoq | 1890 | patah | 2523 | waran |
| 625 | bubuk | 1258 | jongkor | 1891 | pataq | 2524 | waras |
| 626 | bubun | 1259 | jonjoq | 1892 | paten | 2525 | wasiat |
| 627 | bubus | 1260 | joraq | 1893 | kepaten | 2526 | wayahan |
| 628 | bucuq | 1261 | joret | 1894 | patih | 2527 | wong |
| 629 | kebucuk | 1262 | jot | 1895 | patiq | 2528 | yakin |
| 630 | budal | 1263 | jotong | 1896 | patuh | 2529 | ye |
| 631 | budun | 1264 | jowong | 1897 | patut | 2530 | zalim |
| 632 | buit | 1265 | juang | 1898 | pawon | 2531 | zaman |
| 633 | buk-buk | 1266 | juaq | 1899 | payah | 2532 | zohor |

Tabel 4.2 Hasil Stemming Kombinasi Nazief-Adriani dan Porter Stemmer

Beserta Jaro-Winkler Distance

| Word | Found on dictionary | Stem word | Found by similarity | Similarity distance | Similar word | SUCCESS |
|-----------|---------------------|-----------|---------------------|---------------------|--------------|---------|
| leq | false | leq | true | 0.93 | laeq | 0 |
| sopoq | true | sopoq | false | 0.0 | | 1 |
| jelo | false | jelo | true | 0.96 | jeloh | 0 |
| dua | false | dua | true | 0.89 | dugaq | 0 |
| kanak | true | kanak | false | 0.0 | | 1 |
| siq | true | siq | false | 0.0 | | 1 |
| besemeton | false | semeton | true | 1.0 | semeton | 1 |
| sino | false | sino | false | 0.0 | | 1 |
| tekocapan | false | tekocapan | false | 0.0 | | 0 |
| kelapahan | false | kelapahan | true | 0.93 | kelampan | 0 |
| lengan | false | lengan | true | 0.9 | langan | 0 |
| jelo | false | jelo | true | 0.96 | jeloh | 0 |
| siq | true | siq | false | 0.0 | | 1 |
| rubin | true | rubin | false | 0.0 | | 1 |

| | | | | | | |
|-------------------|-------|------------|-------|------|-------|---|
| <i>laiq</i> | true | laiq | false | 0.0 | | 1 |
| <i>ndeqna</i> | false | ndeqna | false | 0.0 | | 0 |
| <i>araq</i> | true | araq | false | 0.0 | | 1 |
| <i>reseki</i> | false | reseki | false | 0.0 | | 1 |
| <i>siq</i> | true | siq | false | 0.0 | | 1 |
| <i>uah</i> | true | uah | false | 0.0 | | 1 |
| <i>tama</i> | false | tama | true | 0.96 | tamaq | 1 |
| <i>tipaq</i> | true | tipaq | false | 0.0 | | 1 |
| <i>tianna</i> | false | tian | true | 1.0 | tian | 1 |
| <i>nyekena</i> | false | nyekena | false | 0.0 | | 0 |
| <i>gati</i> | true | gati | false | 0.0 | | 1 |
| <i>siq</i> | true | siq | false | 0.0 | | 1 |
| <i>napak</i> | false | napak | true | 0.93 | apak | 1 |
| <i>pesilayan</i> | false | pesilayan | false | 0.0 | | 0 |
| <i>amaqna</i> | false | amaq | true | 1.0 | amaq | 1 |
| <i>tesatuq</i> | false | tesatuq | true | 0.89 | tetu | 0 |
| <i>isiq</i> | true | isiq | false | 0.0 | | 1 |
| <i>inaq</i> | true | inaq | false | 0.0 | | 1 |
| <i>tereqna</i> | false | tereqna | false | 0.0 | | 0 |
| <i>ngelor</i> | false | lor | true | 0.93 | leor | 0 |
| <i>tengari</i> | true | tengari | false | 0.0 | | 1 |
| <i>kanak</i> | true | kanak | false | 0.0 | | 1 |
| <i>siq</i> | true | siq | false | 0.0 | | 1 |
| <i>beciqan</i> | false | beciqan | true | 0.95 | beciq | 1 |
| <i>mulai</i> | false | mulai | false | 0.0 | | 1 |
| <i>nangis</i> | false | nangis | true | 0.92 | nani | 0 |
| <i>rue</i> | false | rue | true | 1.0 | rue | 1 |
| <i>teragaqna</i> | false | teragaqna | false | 0.0 | | 0 |
| <i>kanak</i> | true | kanak | false | 0.0 | | 1 |
| <i>siq</i> | true | siq | false | 0.0 | | 1 |
| <i>becig</i> | false | cig | false | 0.0 | | 1 |
| <i>sino</i> | false | sino | false | 0.0 | | 1 |
| <i>mulena</i> | false | mule | true | 1.0 | mule | 1 |
| <i>lapah</i> | true | lapah | false | 0.0 | | 1 |
| <i>rapetangna</i> | false | rapetangna | true | 0.92 | rapet | 1 |
| <i>diriqna</i> | false | diriqna | false | 0.0 | | 0 |
| <i>isiq</i> | true | isiq | false | 0.0 | | 1 |
| <i>kakaq</i> | true | kakaq | false | 0.0 | | 1 |
| <i>siq</i> | true | siq | false | 0.0 | | 1 |
| <i>ngumba</i> | false | ngumba | false | 0.0 | | 1 |
| <i>adiqna</i> | false | adiq | true | 1.0 | adiq | 1 |
| <i>sino</i> | false | sino | false | 0.0 | | 1 |
| <i>adeqna</i> | false | adeq | true | 1.0 | adeq | 1 |
| <i>dengah</i> | true | dengah | false | 0.0 | | 1 |

| | | | | | | |
|-------------------|-------|-----------|-------|------|-------|---|
| <i>isiq</i> | true | isiq | false | 0.0 | | 1 |
| <i>amaqna</i> | false | amaq | true | 1.0 | amaq | 1 |
| <i>suare</i> | false | suare | false | 0.0 | | 1 |
| <i>tangis</i> | true | tangis | false | 0.0 | | 1 |
| <i>adiqna</i> | false | adiq | true | 1.0 | adiq | 1 |
| <i>siq</i> | true | siq | false | 0.0 | | 1 |
| <i>lapar</i> | false | lapar | true | 0.92 | lapah | 1 |
| <i>laguq</i> | false | laguq | false | 0.0 | | 1 |
| <i>siq</i> | true | siq | false | 0.0 | | 1 |
| <i>antihna</i> | false | antih | true | 1.0 | antih | 1 |
| <i>endeqna</i> | false | endeq | true | 1.0 | endeq | 1 |
| <i>araq</i> | true | araq | false | 0.0 | | 1 |
| <i>beketuan</i> | false | beketuan | true | 0.94 | beket | 0 |
| <i>endeqna</i> | false | endeq | true | 1.0 | endeq | 1 |
| <i>araq</i> | true | araq | false | 0.0 | | 1 |
| <i>nyapaqan</i> | false | nyapaqan | false | 0.0 | | 0 |
| <i>endeqna</i> | false | endeq | true | 1.0 | endeq | 1 |
| <i>araq</i> | true | araq | false | 0.0 | | 1 |
| <i>ngewangsit</i> | false | wangsit | false | 0.0 | | 1 |
| <i>apaapa</i> | false | apaapa | false | 0.0 | | 0 |
| <i>ruaruana</i> | false | ruaruana | false | 0.0 | | 0 |
| <i>amaq</i> | true | amaq | false | 0.0 | | 1 |
| <i>sino</i> | false | sino | false | 0.0 | | 1 |
| <i>nyekana</i> | false | nyekana | false | 0.0 | | 0 |
| <i>pacu</i> | true | pacu | false | 0.0 | | 1 |
| <i>gati</i> | true | gati | false | 0.0 | | 1 |
| <i>besuap</i> | false | suap | true | 1.0 | suap | 1 |
| <i>ketuanna</i> | false | ketuanna | false | 0.0 | | 0 |
| <i>seninaqna</i> | false | seninaqna | false | 0.0 | | 0 |
| <i>kumbeqna</i> | false | kumbeqna | false | 0.0 | | 0 |
| <i>kanaq</i> | false | kanaq | true | 0.92 | kanak | 1 |
| <i>sino</i> | false | sino | false | 0.0 | | 1 |
| <i>nangis</i> | false | nangis | true | 0.92 | nani | 0 |
| <i>doang</i> | false | doang | true | 0.89 | doyan | 0 |
| <i>laguq</i> | false | laguq | false | 0.0 | | 1 |
| <i>inaq</i> | true | inaq | false | 0.0 | | 1 |
| <i>tereq</i> | false | tereq | true | 0.91 | teriq | 0 |
| <i>sino</i> | false | sino | false | 0.0 | | 1 |
| <i>bejawap</i> | false | jawab | true | 1.0 | jawab | 1 |
| <i>endeqna</i> | false | endeq | true | 1.0 | endeq | 1 |
| <i>araq</i> | true | araq | false | 0.0 | | 1 |
| <i>maraqna</i> | false | maraq | true | 1.0 | maraq | 1 |

Tabel 4.3 Hasil Stemming Kombinasi Nazief-Adriani dan Porter Stemmer
Tanpa Jaro-Winkler Distance (Lanjut di Lampiran)

| <i>Word</i> | Found on dictionary | Stem word | Found by similarity | Similarity distance | SUCCESS |
|------------------|----------------------------|------------------|----------------------------|----------------------------|----------------|
| <i>leq</i> | false | leq | false | 0.0 | 1 |
| <i>sopoq</i> | true | sopoq | false | 0.0 | 1 |
| <i>jelo</i> | false | jelo | false | 0.0 | 1 |
| <i>dua</i> | false | dua | false | 0.0 | 1 |
| <i>kanak</i> | true | kanak | false | 0.0 | 1 |
| <i>siq</i> | true | siq | false | 0.0 | 1 |
| <i>beseleton</i> | false | seleton | false | 0.0 | 1 |
| <i>sino</i> | false | sino | false | 0.0 | 1 |
| <i>tekocapan</i> | false | tekocapan | false | 0.0 | 0 |
| <i>kelapahan</i> | false | kelapahan | false | 0.0 | 0 |
| <i>lengan</i> | false | lengan | false | 0.0 | 0 |
| <i>jelo</i> | false | jelo | false | 0.0 | 1 |
| <i>siq</i> | true | siq | false | 0.0 | 1 |
| <i>rubin</i> | true | rubin | false | 0.0 | 1 |
| <i>laiq</i> | true | laiq | false | 0.0 | 1 |
| <i>ndeqna</i> | false | ndeqna | false | 0.0 | 0 |
| <i>araq</i> | true | araq | false | 0.0 | 1 |
| <i>reseki</i> | false | reseki | false | 0.0 | 1 |
| <i>siq</i> | true | siq | false | 0.0 | 1 |
| <i>uah</i> | true | uah | false | 0.0 | 1 |
| <i>tama</i> | false | tama | false | 0.0 | 1 |
| <i>tipaq</i> | true | tipaq | false | 0.0 | 1 |
| <i>tianna</i> | false | tian | false | 0.0 | 1 |
| <i>nyekena</i> | false | nyekena | false | 0.0 | 0 |
| <i>gati</i> | true | gati | false | 0.0 | 1 |
| <i>siq</i> | true | siq | false | 0.0 | 1 |
| <i>napak</i> | false | napak | false | 0.0 | 1 |
| <i>pesilayan</i> | false | pesilayan | false | 0.0 | 0 |
| <i>amaqna</i> | false | amaq | false | 0.0 | 0 |
| <i>tesatuq</i> | false | tesatuq | false | 0.0 | 0 |
| <i>isiq</i> | true | isiq | false | 0.0 | 1 |
| <i>inaq</i> | true | inaq | false | 0.0 | 1 |
| <i>tereqna</i> | false | tereqna | false | 0.0 | 0 |
| <i>ngelor</i> | false | lor | false | 0.0 | 0 |
| <i>tengari</i> | true | tengari | false | 0.0 | 1 |
| <i>kanak</i> | true | kanak | false | 0.0 | 1 |
| <i>siq</i> | true | siq | false | 0.0 | 1 |
| <i>beciqan</i> | false | beciqan | false | 0.0 | 0 |
| <i>mulai</i> | false | mulai | false | 0.0 | 1 |

| | | | | | |
|-------------------|-------|------------|-------|-----|---|
| <i>nangis</i> | false | nangis | false | 0.0 | 0 |
| <i>ruen</i> | false | rue | false | 0.0 | 1 |
| <i>teragaqna</i> | false | teragaqna | false | 0.0 | 0 |
| <i>kanak</i> | true | kanak | false | 0.0 | 1 |
| <i>siq</i> | true | siq | false | 0.0 | 1 |
| <i>becig</i> | false | cig | false | 0.0 | 0 |
| <i>sino</i> | false | sino | false | 0.0 | 1 |
| <i>mulena</i> | false | mule | false | 0.0 | 1 |
| <i>lapah</i> | true | lapah | false | 0.0 | 1 |
| <i>rapetangna</i> | false | rapetangna | false | 0.0 | 0 |
| <i>diriqna</i> | false | diriqna | false | 0.0 | 0 |
| <i>isiq</i> | true | isiq | false | 0.0 | 1 |
| <i>kakaq</i> | true | kakaq | false | 0.0 | 1 |
| <i>siq</i> | true | siq | false | 0.0 | 1 |
| <i>ngumbaq</i> | false | ngumbaq | false | 0.0 | 0 |
| <i>adiqna</i> | false | adiq | false | 0.0 | 1 |
| <i>sino</i> | false | sino | false | 0.0 | 1 |
| <i>adeqna</i> | false | adeq | false | 0.0 | 1 |
| <i>dengah</i> | true | dengah | false | 0.0 | 1 |
| <i>isiq</i> | true | isiq | false | 0.0 | 1 |
| <i>amaqna</i> | false | amaq | false | 0.0 | 1 |
| <i>suare</i> | false | suare | false | 0.0 | 1 |
| <i>tangis</i> | true | tangis | false | 0.0 | 1 |
| <i>adiqna</i> | false | adiq | false | 0.0 | 1 |
| <i>siq</i> | true | siq | false | 0.0 | 1 |
| <i>lapar</i> | false | lapar | false | 0.0 | 0 |
| <i>laguq</i> | false | laguq | false | 0.0 | 1 |
| <i>siq</i> | true | siq | false | 0.0 | 1 |
| <i>antihna</i> | false | antih | false | 0.0 | 1 |
| <i>endeqna</i> | false | endeq | false | 0.0 | 1 |
| <i>araq</i> | true | araq | false | 0.0 | 1 |
| <i>beketuan</i> | false | beketuan | false | 0.0 | 0 |
| <i>endeqna</i> | false | endeq | false | 0.0 | 1 |
| <i>araq</i> | true | araq | false | 0.0 | 1 |
| <i>nyapaqan</i> | false | nyapaqan | false | 0.0 | 0 |
| <i>endeqna</i> | false | endeq | false | 0.0 | 1 |
| <i>araq</i> | true | araq | false | 0.0 | 1 |
| <i>ngewangsit</i> | false | wangsit | false | 0.0 | 1 |
| <i>apaapa</i> | false | apaapa | false | 0.0 | 0 |
| <i>ruaruana</i> | false | ruaruana | false | 0.0 | 0 |
| <i>amaq</i> | true | amaq | false | 0.0 | 1 |
| <i>sino</i> | false | sino | false | 0.0 | 1 |
| <i>nyekana</i> | false | nyekana | false | 0.0 | 0 |
| <i>pacu</i> | true | pacu | false | 0.0 | 1 |

| | | | | | |
|------------------|-------|------------------|-------|-----|---|
| <i>gati</i> | true | <i>gati</i> | false | 0.0 | 1 |
| <i>besuap</i> | false | <i>suap</i> | false | 0.0 | 1 |
| <i>ketuanna</i> | false | <i>ketuanna</i> | false | 0.0 | 0 |
| <i>seninaqna</i> | false | <i>seninaqna</i> | false | 0.0 | 0 |
| <i>kumbeqna</i> | false | <i>kumbeqna</i> | false | 0.0 | 0 |
| <i>kanaq</i> | false | <i>kanaq</i> | false | 0.0 | 1 |
| <i>sino</i> | false | <i>sino</i> | false | 0.0 | 1 |
| <i>nangis</i> | false | <i>nangis</i> | false | 0.0 | 1 |
| <i>doang</i> | false | <i>doang</i> | false | 0.0 | 1 |
| <i>laguq</i> | false | <i>laguq</i> | false | 0.0 | 1 |
| <i>inaq</i> | true | <i>inaq</i> | false | 0.0 | 1 |
| <i>tereq</i> | false | <i>tereq</i> | false | 0.0 | 1 |
| <i>sino</i> | false | <i>sino</i> | false | 0.0 | 1 |
| <i>bejawab</i> | false | <i>jawab</i> | false | 0.0 | 1 |
| <i>endeqna</i> | false | <i>endeqna</i> | false | 0.0 | 1 |
| <i>araq</i> | true | <i>Araq</i> | false | 0.0 | 1 |
| <i>maraqna</i> | false | <i>Maraq</i> | false | 0.0 | 1 |

Tabel 4.5 Hasil Stemming Kombinasi Nazief-Adriani dan Porter Stemmer

Tanpa Jaro-Winkler Distance dan Tanpa Dictionary

| Word | Found on dictionary | Stem word | SUCCESS |
|------------------|----------------------------|------------------|----------------|
| <i>leq</i> | false | <i>leq</i> | 1 |
| <i>sopoq</i> | false | <i>sopoq</i> | 1 |
| <i>jelo</i> | false | <i>jelo</i> | 1 |
| <i>dua</i> | false | <i>dua</i> | 1 |
| <i>kanak</i> | false | <i>kanak</i> | 1 |
| <i>siq</i> | false | <i>siq</i> | 1 |
| <i>besemeton</i> | false | <i>semeton</i> | 1 |
| <i>sino</i> | false | <i>sino</i> | 1 |
| <i>tekocapan</i> | false | <i>tekocapan</i> | 0 |
| <i>kelapahan</i> | false | <i>kelapahan</i> | 0 |
| <i>lengan</i> | false | <i>lengan</i> | 1 |
| <i>jelo</i> | false | <i>jelo</i> | 1 |
| <i>siq</i> | false | <i>siq</i> | 1 |
| <i>rubin</i> | false | <i>rubin</i> | 1 |
| <i>laiq</i> | false | <i>laiq</i> | 1 |
| <i>ndeqna</i> | false | <i>ndeqna</i> | 0 |
| <i>araq</i> | false | <i>araq</i> | 1 |
| <i>reseki</i> | false | <i>reseki</i> | 1 |
| <i>siq</i> | false | <i>siq</i> | 1 |
| <i>uah</i> | false | <i>uah</i> | 1 |

| | | | |
|-------------------|-------|------------|---|
| <i>tama</i> | false | tama | 1 |
| <i>tipaq</i> | false | tipaq | 1 |
| <i>tianna</i> | false | tianna | 0 |
| <i>nyekena</i> | false | nyekena | 0 |
| <i>gati</i> | false | gati | 1 |
| <i>siq</i> | false | siq | 1 |
| <i>napak</i> | false | napak | 1 |
| <i>pesilayan</i> | false | pesilayan | 0 |
| <i>amaqna</i> | false | amaqna | 0 |
| <i>tesatuq</i> | false | tesatuq | 0 |
| <i>isiq</i> | false | isiq | 1 |
| <i>inaq</i> | false | inaq | 1 |
| <i>tereqna</i> | false | tereqna | 0 |
| <i>ngelor</i> | false | lor | 0 |
| <i>tengari</i> | false | tengari | 1 |
| <i>kanak</i> | false | kanak | 1 |
| <i>siq</i> | false | siq | 1 |
| <i>beciqan</i> | false | beciqan | 0 |
| <i>mulai</i> | false | mulai | 1 |
| <i>nangis</i> | false | nangis | 0 |
| <i>ruen</i> | false | ruen | 1 |
| <i>teragaqna</i> | false | teragaqna | 0 |
| <i>kanak</i> | false | kanak | 1 |
| <i>siq</i> | false | siq | 1 |
| <i>becig</i> | false | cig | 0 |
| <i>sino</i> | false | sino | 1 |
| <i>mulena</i> | false | mulena | 0 |
| <i>lapah</i> | false | lapah | 1 |
| <i>rapetangna</i> | false | rapetangna | 0 |
| <i>diriqna</i> | false | diriqna | 0 |
| <i>isiq</i> | false | isiq | 1 |
| <i>kakaq</i> | false | kakaq | 1 |
| <i>siq</i> | false | siq | 1 |
| <i>ngumbaq</i> | false | ngumbaq | 1 |
| <i>adiqna</i> | false | adiqna | 0 |
| <i>sino</i> | false | sino | 1 |
| <i>adeqna</i> | false | adeqna | 0 |
| <i>dengah</i> | false | dengah | 1 |
| <i>isiq</i> | false | isiq | 1 |
| <i>amaqna</i> | false | amaqna | 0 |
| <i>suare</i> | false | suare | 1 |
| <i>tangis</i> | false | tangis | 1 |
| <i>adiqna</i> | false | adiqna | 0 |
| <i>siq</i> | false | siq | 1 |

| | | | |
|-------------------|-------|-----------|---|
| <i>lapar</i> | false | lapar | 0 |
| <i>laguq</i> | false | laguq | 1 |
| <i>siq</i> | false | siq | 1 |
| <i>antihna</i> | false | antihna | 0 |
| <i>endeqna</i> | false | endeqna | 0 |
| <i>araq</i> | false | araq | 1 |
| <i>beketuan</i> | false | beketuan | 0 |
| <i>endeqna</i> | false | endeqna | 0 |
| <i>araq</i> | false | araq | 1 |
| <i>nyapaqan</i> | false | nyapaqan | 0 |
| <i>endeqna</i> | false | endeqna | 0 |
| <i>araq</i> | false | araq | 1 |
| <i>ngewangsit</i> | false | wangsit | 0 |
| <i>apaapa</i> | false | apaapa | 0 |
| <i>ruaruana</i> | false | ruaruana | 0 |
| <i>amaq</i> | false | amaq | 1 |
| <i>sino</i> | false | sino | 1 |
| <i>nyekana</i> | false | nyekana | 0 |
| <i>pacu</i> | false | pacu | 1 |
| <i>gati</i> | false | gati | 1 |
| <i>besuap</i> | false | suap | 1 |
| <i>ketuanna</i> | false | ketuanna | 0 |
| <i>seninaqna</i> | false | seninaqna | 0 |
| <i>kumbegna</i> | false | kumbegna | 0 |
| <i>kanaq</i> | false | kanaq | 1 |
| <i>sino</i> | false | sino | 1 |
| <i>nangis</i> | false | nangis | 0 |
| <i>doang</i> | false | doang | 1 |
| <i>laguq</i> | false | laguq | 1 |
| <i>inaq</i> | false | inaq | 1 |
| <i>tereq</i> | false | tereq | 1 |
| <i>sino</i> | false | sino | 1 |
| <i>bejawab</i> | false | jawab | 1 |
| <i>endeqna</i> | false | endeqna | 0 |
| <i>araq</i> | False | araq | 1 |
| <i>maraqna</i> | False | maraqna | 0 |

Tabel 4.6 Hasil Stemming Nazief-Adriani Tanpa Jaro-Winkler Distance

| Word | Found on dictionary | Stem word | SUCCESS |
|--------------|----------------------------|------------------|----------------|
| <i>leq</i> | false | leq | 1 |
| <i>sopoq</i> | true | sopoq | 1 |
| <i>jelo</i> | false | jelo | 1 |

| | | | |
|------------------|-------|-----------|---|
| <i>dua</i> | false | dua | 1 |
| <i>kanak</i> | true | kanak | 1 |
| <i>siq</i> | true | siq | 1 |
| <i>besemeton</i> | false | besemeton | 0 |
| <i>sino</i> | false | sino | 1 |
| <i>tekocapan</i> | false | tekocapan | 0 |
| <i>kelapahan</i> | false | kelapahan | 0 |
| <i>lengan</i> | false | lengan | 1 |
| <i>jelo</i> | false | jelo | 1 |
| <i>siq</i> | true | siq | 1 |
| <i>rubin</i> | true | rubin | 1 |
| <i>laiq</i> | true | laiq | 1 |
| <i>ndeqna</i> | false | ndeqna | 1 |
| <i>araq</i> | true | araq | 1 |
| <i>reseki</i> | false | reseki | 1 |
| <i>siq</i> | true | siq | 1 |
| <i>uah</i> | true | uah | 1 |
| <i>tama</i> | false | tama | 1 |
| <i>tipaq</i> | true | tipaq | 1 |
| <i>tianna</i> | false | tianna | 0 |
| <i>nyekena</i> | false | nyekena | 0 |
| <i>gati</i> | true | gati | 1 |
| <i>siq</i> | true | siq | 1 |
| <i>napak</i> | false | napak | 1 |
| <i>pesilayan</i> | false | pesilayan | 0 |
| <i>amaqna</i> | false | amaqna | 0 |
| <i>tesatuq</i> | false | tesatuq | 0 |
| <i>isiq</i> | true | isiq | 1 |
| <i>inaq</i> | true | inaq | 1 |
| <i>tereqna</i> | false | tereqna | 0 |
| <i>ngelor</i> | false | ngelor | 1 |
| <i>tengari</i> | true | tengari | 1 |
| <i>kanak</i> | true | kanak | 1 |
| <i>siq</i> | true | siq | 1 |
| <i>beciqan</i> | false | beciqan | 0 |
| <i>mulai</i> | false | mulai | 1 |
| <i>nangis</i> | false | nangis | 0 |
| <i>ruen</i> | false | ruen | 1 |
| <i>teragaqna</i> | false | teragaqna | 0 |
| <i>kanak</i> | true | kanak | 1 |
| <i>siq</i> | true | siq | 1 |
| <i>becig</i> | false | becig | 1 |
| <i>sino</i> | false | sino | 1 |
| <i>mulena</i> | false | mulena | 0 |

| | | | |
|-------------------|-------|------------|---|
| <i>lapah</i> | true | lapah | 1 |
| <i>rapetangna</i> | false | rapetangna | 0 |
| <i>diriqna</i> | false | diriqna | 0 |
| <i>isiq</i> | true | isiq | 1 |
| <i>kakaq</i> | true | kakaq | 1 |
| <i>siq</i> | true | siq | 1 |
| <i>ngumbaq</i> | false | ngumbaq | 1 |
| <i>adiqna</i> | false | adiqna | 0 |
| <i>sino</i> | false | sino | 1 |
| <i>adeqna</i> | false | adeqna | 0 |
| <i>dengah</i> | true | dengah | 1 |
| <i>isiq</i> | true | isiq | 1 |
| <i>amaqna</i> | false | amaqna | 0 |
| <i>suare</i> | false | suare | 1 |
| <i>tangis</i> | true | tangis | 1 |
| <i>adiqna</i> | false | adiqna | 0 |
| <i>siq</i> | true | siq | 1 |
| <i>lapar</i> | false | lapar | 0 |
| <i>laguq</i> | false | laguq | 1 |
| <i>siq</i> | true | siq | 1 |
| <i>antihna</i> | false | antihna | 0 |
| <i>endeqna</i> | false | endeqna | 0 |
| <i>araq</i> | true | araq | 1 |
| <i>beketuan</i> | false | beketuan | 0 |
| <i>endeqna</i> | false | endeqna | 0 |
| <i>araq</i> | true | araq | 1 |
| <i>nyapaqan</i> | false | nyapaqan | 0 |
| <i>endeqna</i> | false | endeqna | 0 |
| <i>araq</i> | true | araq | 1 |
| <i>ngewangsit</i> | false | ngewangsit | 0 |
| <i>apaapa</i> | false | apaapa | 0 |
| <i>ruaruana</i> | false | ruaruana | 0 |
| <i>amaq</i> | true | amaq | 1 |
| <i>sino</i> | false | sino | 1 |
| <i>nyekana</i> | false | nyekana | 0 |
| <i>pacu</i> | true | pacu | 1 |
| <i>gati</i> | true | gati | 1 |
| <i>besuap</i> | false | besuap | 0 |
| <i>ketuanna</i> | false | ketuanna | 0 |
| <i>seninaqna</i> | false | seninaqna | 0 |
| <i>kumbeqna</i> | false | kumbeqna | 0 |
| <i>kanaq</i> | false | kanaq | 1 |
| <i>sino</i> | false | sino | 1 |
| <i>nangis</i> | false | nangis | 0 |

| | | | |
|----------------|-------|---------|---|
| <i>doang</i> | false | doang | 1 |
| <i>laguq</i> | false | laguq | 1 |
| <i>inaq</i> | true | inaq | 1 |
| <i>terek</i> | false | terek | 1 |
| <i>sino</i> | false | sino | 1 |
| <i>bejawab</i> | false | bejawab | 0 |
| <i>endeqna</i> | false | endeqna | 0 |
| <i>araq</i> | true | araq | 1 |
| <i>maraqna</i> | false | maraqna | 0 |

