

## **Mapping and Analysis of Standard Indonesian Pronunciation Errors by Using the Bigram Method**

(Center, Bold, Cambria 12, Capitalized)

**Emmy Erwina<sup>1</sup>, Tommy<sup>2</sup>, Mayasari<sup>3</sup>**

Universitas Harapan Medan<sup>1,2,3</sup>, Medan, Indonesia

[tomshirakawa@gmail.com](mailto:tomshirakawa@gmail.com)

**Abstract:** Indonesian language is increasingly being ignored, even the mass media often find the use of non-standard language, so there is a uniformity in the use of words that often appear in scientific articles, especially those Indonesian. The uniformity of Indonesian pronunciation certainly confuses the general public, for example: television news viewers and radio listeners, to distinguish between standard and non-standard forms. The non-uniformity of Indonesian pronunciation often occurs in official situations such as official speeches or presentations. Based on this phenomenon, this study aims to conduct an analysis and mapping study of standard pronunciation errors that arise in several public services and develops application tools as automatically tool that can detect errors from the use of standard words obtained from voice recording results in the form of vowel errors, diphthongs and consonants by implementing the bigram method. This research was conducted by collecting voice recordings of conversations, speeches, public speaking and other recordings deemed necessary, then continued with standard pronunciation errors analysis and mapping of these errors using a tool that applied the Bigram technique. Based on statistical information obtained from the detection and correction process that has been carried out, it can be obtained a mapping of the error types in the use of Indonesian standard that often occur in public service environment. The results of the mapping and statistical information obtained are used as the basis for developing programs that aim to improve the quality of the use Indonesian standard in the form of FGDs, socialization and counseling. The study result indicate the standard of pronunciation errors that occur in various circles including presenters and lecturers, then the results of data analysis are described in the form of tables and diagrams so it is seen the percentage of errors that occur.

**Keywords:** Language Error; Standard Language; Bigram; Presenter; Lectures

### **INTRODUCTION**

Detection and mapping of standard language errors are very important in the modern era like today where awareness of using good language is very low where there is high use of market or non-standard languages in daily activities there are various regional languages or dialects, then in its development Indonesian has always been influenced by regional languages or dialects, such as Minangkabau, Javanese, Sundanese, and Batak languages. In addition, Indonesian is also influenced by foreign languages, such as Dutch, English, Arabic and Sanskrit (Badudu, 1993:63).

The influence of foreign languages, regional languages or dialects on Indonesian can be seen in pronunciation or pronunciation. This influence causes a non-uniformity of pronunciation, this may be the reason why until now in Indonesia there has not been a



standard pronunciation guideline for Indonesian (Erwina, 2012: 1). The non-uniformity of Indonesian pronunciation certainly confuses the general public, for example news viewers on television, radio listeners and official meetings, to distinguish the standard form and the non-standard form. Fishman (1972:17) says that official means a regulation that has been codified, accepted in the user community and supported by the government, the world of education, the mass media and others.

The Center for Language Development and Development (now called the Language and Book Development Agency) has not yet issued guidelines for Indonesian pronunciation or pronunciation (Badudu, 1995:206). However, the Center for Language Development and Development has issued General Guidelines for Indonesian Spelling Enhanced (since 2015 PUEBIYD has changed to General Guidelines for Indonesian Spelling) and General Guidelines for the Formation of Terms and a Standard Indonesian Grammar book.

With the two guidebooks, it can be said that the standard pronunciation of Indonesian is the pronunciation that follows the phonemic principle based on Indonesian spelling, meaning that a word is pronounced following the standard Indonesian spelling or according to the sound value of the letters in Indonesian. For example, the word 'eat' is pronounced as [eat] and the word 'problem' is pronounced as [problem]. (Erwina, 2012:2). According to Singgih (1972:72), however, the Indonesian pronunciation needs to be standardized, for example the sound described with the letter 'a', wherever its place in the word or pronunciation must be referred to as [a] and not changed to [e], [ə], or [ɛ].

Many studies related to the use of standard language errors have also been carried out, one of which is that not all students can write scientific papers using good and correct Indonesian (Jamilah, 2017). Other research states that at a school the students have a fairly high rate of language use errors (Tribana, 2012). Based on these problems, this study aims to analyze and map errors from standard words obtained from sound recordings obtained from public services such as educational institutions, media, and other environments that become community guidelines in everyday language. This study will analyzed standard language errors in the form of vowel errors, diphthongs and consonants. The error detection carried out in this study uses the Bigram technique which is applied to an application system that can increase speed and accuracy in the statistical process and mapping errors in the use of standard Indonesian pronunciation so that good and accountable mapping information can be produced against the phenomenon of the decline in the use of standard pronunciation of the language. Indonesia is good at this time.

## METHOD

Standard words are the words whose pronunciation or writing is under standard rules or rules that have been standardized. The variety of standard words has dynamic stability, meaning that it has rules and rules in language standards. The standard word is usually used in official situations, such as seminars, speeches, scientific work gatherings, etc. The standard rules in question can be in the form of General Guidelines for Indonesian Spelling

115

Publish By

The Association of Community Service Lecturers of Indonesia (ADPI)

Jl. Komp. Unand Padang Besi Kec. Lubuk Kilangan, Kota Padang. Sumatera Barat

Code 25166

Website: [www.adpi-indonesia.id](http://www.adpi-indonesia.id)Email : [info@adpi-indonesia.id](mailto:info@adpi-indonesia.id)

Phone : +62 823-8662-2140



(EYD), general dictionaries, and standard grammar. In this case, the Indonesian dictionary can be used as a reference to find out the standard language. The emergence of the term standard words or standard words can be traced from the history of language development which shows that various educated people acquire high prestige and authority. This variety is then used as a benchmark for the correct use of language. The error that often occurs is that the standardization of words has not been socialized to the community. Thus, people do not know which spellings, words, or sentence structures are standard and which are not. People often do not have the opportunity to see the rules of the standard. They often hear people talk, or read articles in magazines or other print media. Meanwhile, those who speak or write also of course use standard things.

**N-Gram Language Model** Conceptually, the n-gram model is an estimate of the probability of a word or character from a history of previous occurrences (Jurafsky & Martin, 2018). Mathematically the n-gram model looks for occurrences of n-1 words or characters from the previous words series. N-gram itself consists of several types such as bigram which looks for the appearance of a pair of characters, trigram which looks for the appearance of three pairs of characters and other gram forms. The assumption of the probability of the emergence of a word or character that depends on the previous words is called the Markov assumption (Jurafsky & Martin, 2018).

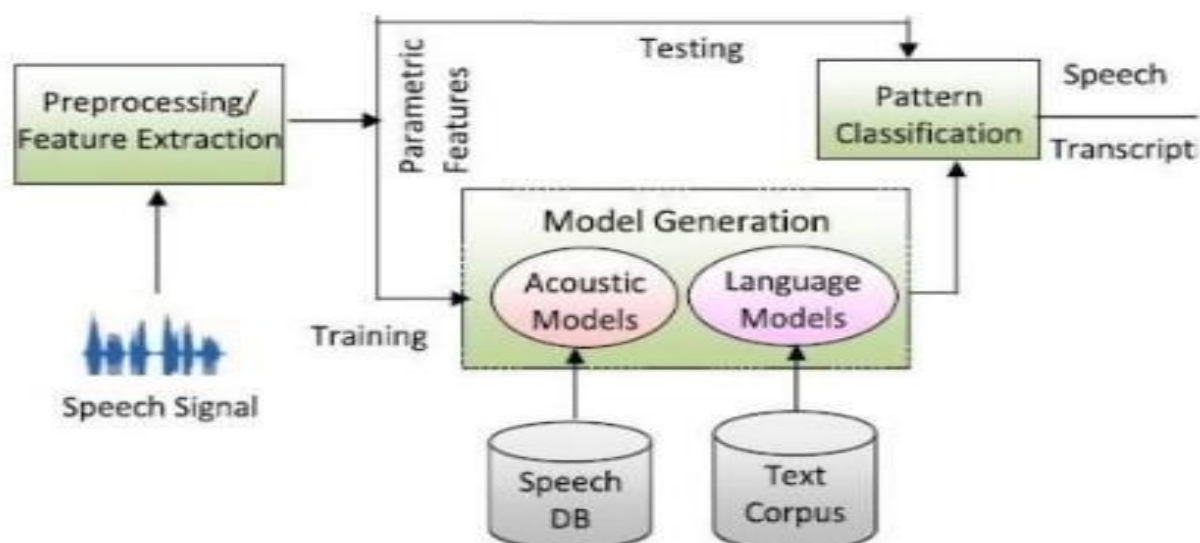
A bigram is a language detection model which is one of the most widely applied n-gram concepts in the field of Natural Language Processing (NLP). Like n-grams, bigrams look for the probability value of the occurrence of a character or word using the history of previous occurrences or using the corpus.

To calculate the probability of the gram model used by dividing the frequency of words used from a certain order by the frequency based on the prefix which can be used to measure the value of MLE. Where a good MLE indicates the suitability of the corpus used. Bigram itself has been widely applied in several studies, especially in the field of NLP. Some implementations can be in the form of word classification (Fahrudin, Buliali, & Fatichah, 2019) and word error detection (Samanta & Chaudhuri, 2013).

Speech Recognition, which is also known as Automatic Speech Recognition (ASR) is a conversion process from voice signals (audio) into word sequences using an algorithm that is implemented in a computer program (Anusuya & Katti, 2009). Speech Recognition can be divided into several components, namely (Trivedi, Pant, Shah, Sonik, & Agrawal, 2018):

1. Speaker, sound object or sound identity.
2. Vocal Sound, aspects of the way speakers speak (high, low, pause, etc.)
3. Vocabulary, complexity and precision of the size of the vocabulary used.

The main stages of speech recognition consist of several main stages, namely pre-processing, feature extraction, acoustic models, language models, pattern classification and several other additional stages. Some studies use a slightly different approach which is very dependent on the method or technique used.



**Figure 1. Speech Recognition System Architecture**  
(Source: Trivedi, Pant, Shah, Sonik, & Agrawal, 2018)

The research method used in this study uses a qualitative and quantitative approach. The data collection used in this study used a case study method where data was collected from interviews, speaker/speech manuscript collection, voice recordings of presentations and other documents deemed necessary obtained from educational institutions and mass media stations so that the data collected is expected to represent the phenomenon of use. Inappropriate standard language in the educational environment and information media.

The population used in this study consists of two groups, where the first group is academics who can be teachers or lecturers, while the second group is broadcasters at mass media stations. The detailed population groups used can be described as follows:

1. Educators, which consist of educators from elementary, junior high, high school levels and lecturers who are still active in teaching and learning activities.
2. TVRI broadcasters, broadcasters classified as active in broadcasting activities at TVRI stations in Sumatera Utara.

The selection of the population or respondents above is based on the fact that the two population groups are active in conveying information to other individuals so that they become guidelines for other individuals, both students and the community in language.

The types of data used in this study are data in the form of sound documents or texts obtained from the population mentioned above. Data were obtained through several approaches such as recording live interviews, recording broadcasts, recording during exposure activities, especially for academics. The recording of direct interviews and



presentations of teaching and learning activities is carried out using a recording device which can then be further analyzed during the analysis and mapping process.

The description of the stages of this research are:

1. Study of phenomena and literature, at this stage the Chief Researcher and Research Member 2 jointly observe and observe the phenomenon of the high level of errors in the use of standard language both from general observations and from the previous research literature.
2. Identification of problems, at this stage the Chief Researcher formulates problems that arise from the phenomenon that occurs, namely the high level of errors in the use of standard language so that it is necessary to map and classify the classes of standard language errors so that appropriate solutions can be applied to help improve the quality of language use. raw.
3. Primary and Secondary Data Collection, at this stage data collection is carried out in the form of voice recordings, scripts and other data relating to the use of the Indonesian language in the form of communication to the community obtained from two respondents, namely educators in educational institutions and broadcasters from mass media by the lead researcher and research members 2. Supporting data is carried out by compiling a Corpus of standard and non-standard language from the literature and questionnaires conducted by research members 1.
4. Normalization of Data, at this stage normalization of data is carried out so that the data used can be further processed using the application tools used. Voice recording data or scripts will be processed first, such as removing noise contained in sound recordings, converting physical documents into digital format and other forms of normalization carried out by research member 1 who is evaluated by the lead researcher and research member 2.
5. Standard Error Detection, at this stage any normalized data will be processed using tools built by applying the bigram method to detect standard language errors contained in the data carried out by research member 1 who is accompanied by the lead researcher and research member 2 for evaluation the results of the activity stages.
6. Statistical analysis of standard errors, at this stage the frequency and type of errors contained in the processed data will then be further processed to obtain statistical patterns of each type of standard use error. The lead researcher here is in charge of coordinating and evaluating statistical analysis activities carried out by research members 1 and research members 2.
7. Classification and Mapping of Standard Errors in the Language, at this stage, classification is carried out into several groups to obtain information on the mapping of standard errors from educators and broadcasters. At this stage, the lead researcher and research member 2 together carry out classification and mapping of errors in the use of standard language with the help of research member 1 who prepares the tools or applications needed.





8. Discussion of the results of the mapping, at this stage the results of the mapping will be discussed and socialized in the form of FGDs as well as counseling and socialization seminars conducted by the lead researcher as the main resource person accompanied by research member 1 and research member 2 as additional resource persons.

9. Prototype development, at this stage the results of data collection and updating of errors-Standard language errors will be developed and designed into an application prototype that can detect and automatically correct errors in the use of standard language made by research members 1 who are accompanied by the head researcher and research member 2 to provide input and evaluation.

10. Conclusion, at this stage every statistical and mapping information will be formulated into the conclusions of the results of the research carried out jointly by the head researcher and research members.

## RESULTS & DISCUSSION (Cambria 12, bold, Capitalized)

Based on the data analysis that has been done, the research results are obtained as described in the table and diagram below:

### a. Data Group By Presenter

id_data	id_person	id_jenis_kata	id_jenis_salah	id_profesi	id_tingkat_profesi	jk	usia	label_profesi	label_tingkat_profesi	label_jenis	kelas_salah	kat
0	1	1	3	1	1	1 Laki-Laki	46	Guru	SD	Kata Keterangan		Vokal
1	2	1	1	2	1	1 Laki-Laki	46	Guru	SD	Kata Kerja		Diftong se
2	3	1	1	1	1	1 Laki-Laki	46	Guru	SD	Kata Kerja		Vokal
3	4	1	2	1	1	1 Laki-Laki	46	Guru	SD	Kata Sifat		Vokal
4	5	1	9	3	1	1 Laki-Laki	46	Guru	SD	Kata Hubung		Konsonan
...	...	...	...	...	...	...	...	...	...	...	...	...
102	126	12	9	2	2	4 Laki-Laki	35	Dosen	S1	Kata Hubung		Diftong
103	127	12	9	3	2	4 Laki-Laki	35	Dosen	S1	Kata Hubung		Konsonan
104	128	12	1	3	2	4 Laki-Laki	35	Dosen	S1	Kata Kerja		Konsonan
105	129	12	9	2	2	4 Laki-Laki	35	Dosen	S1	Kata Hubung		Diftong
106	130	12	12	2	2	4 Laki-Laki	35	Dosen	S1	Kata Tanya		Diftong ba

107 rows x 13 columns

Notes:

- Usia* : age
- Tingkat* : level
- Kelas* : class
- Jenis* : types
- Salah* : false
- Laki-laki* : male
- Dosen* : lectures
- Kata* : words



*Kata keterangan* : adverbs  
*Kata kerja* : verbs  
*Kata sifat* : adjectives  
*Kata hubung* : conjunctions  
*Kata tanya* : interrogatives  
*Konsonan* : consonant  
*Diftong* : diphthong

id_person	id_jenis_salah	id_profesi	id_tingkat_profesi	jk	usia	label_profesi	label_tingkat_profesi	kelas_salah	kata_salah	
0	1	1	1	1	Laki-Laki	46	Guru	SD	Vokal	12
1	1	2	1	1	Laki-Laki	46	Guru	SD	Diftong	2
2	1	3	1	1	Laki-Laki	46	Guru	SD	Konsonan	3
3	2	1	1	1	Laki-Laki	41	Guru	SD	Vokal	7
4	2	2	1	1	Laki-Laki	41	Guru	SD	Diftong	3
5	2	3	1	1	Laki-Laki	41	Guru	SD	Konsonan	1
6	3	1	1	2	Laki-Laki	44	Guru	SMP	Vokal	2
7	3	2	1	2	Laki-Laki	44	Guru	SMP	Diftong	2
8	3	3	1	2	Laki-Laki	44	Guru	SMP	Konsonan	2
9	4	1	1	2	Perempuan	30	Guru	SMP	Vokal	2
10	4	2	1	2	Perempuan	30	Guru	SMP	Diftong	1
11	4	3	1	2	Perempuan	30	Guru	SMP	Konsonan	3
12	5	1	1	2	Laki-Laki	46	Guru	SMP	Vokal	4
13	5	2	1	2	Laki-Laki	46	Guru	SMP	Diftong	2
14	5	3	1	2	Laki-Laki	46	Guru	SMP	Konsonan	1
15	6	1	1	1	Laki-Laki	40	Guru	SD	Vokal	3
16	6	3	1	1	Laki-Laki	40	Guru	SD	Konsonan	2
17	7	1	1	1	Perempuan	28	Guru	SD	Vokal	4
18	8	1	1	2	Laki-Laki	29	Guru	SMP	Vokal	2
19	9	1	2	4	Perempuan	40	Dosen	S1	Vokal	4
20	9	3	2	4	Perempuan	40	Dosen	S1	Konsonan	1
21	10	1	2	4	Perempuan	30	Dosen	S1	Vokal	19
22	10	2	2	4	Perempuan	30	Dosen	S1	Diftong	3
23	10	3	2	4	Perempuan	30	Dosen	S1	Konsonan	8
24	11	1	1	1	Laki-Laki	40	Guru	SD	Vokal	5
25	11	3	1	1	Laki-Laki	40	Guru	SD	Konsonan	2
26	12	2	2	4	Laki-Laki	35	Dosen	S1	Diftong	3
27	12	3	2	4	Laki-Laki	35	Dosen	S1	Konsonan	4

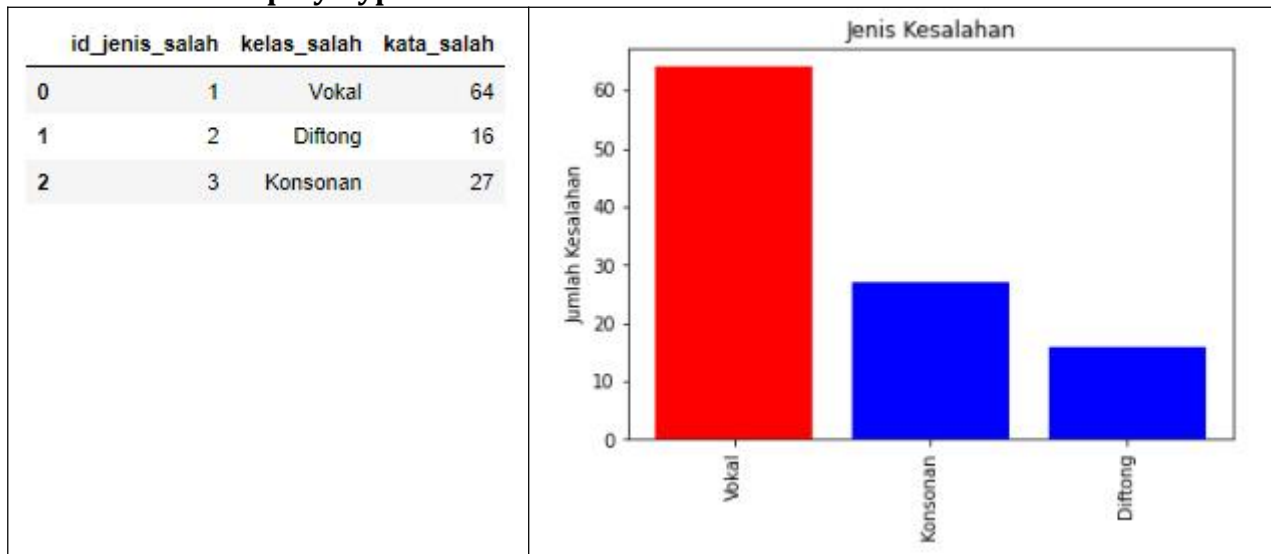
Notes:

*Usia* : age  
*Tingkat* : level  
*Kelas* : class  
*Jenis* : types  
*Salah* : false  
*Laki-laki* : male  
*Perempuan* : female  
*Profesi* : profession  
*Guru* : teachers



*Dosen* : lectures  
*Konsonan* : consonant  
*Diftong* : diphthong

**b. Data Group By Types of Errors**



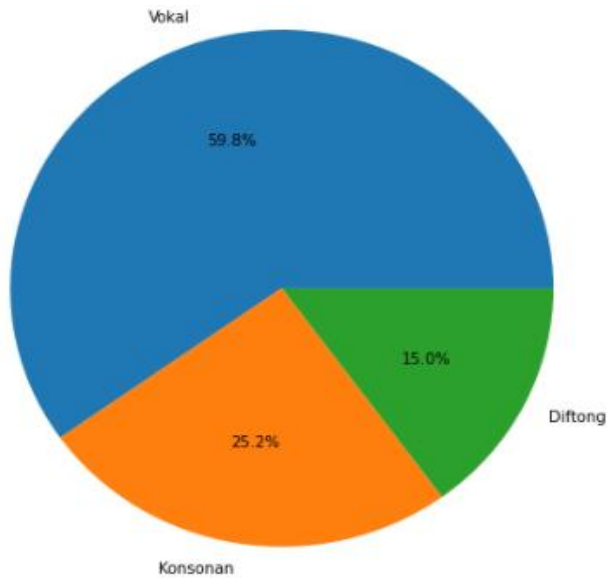
Notes:

*Kelas* : class  
*Jenis* : types  
*Salah* : false  
*Vokal* : vocal  
*Konsonan* : consonant  
*Diftong* : diphthong  
*Jenis kesalahan* : types of errors  
*Jumlah kesalahan* : results of errors





Percentage of errors for each type of errors:



Notes:

*Vokal* : vocal

*Konsonan* : consonant

*Diftong* : diphthong

### c. Data Group By Gender

- The Data of Gender to types of errors

	jk	id_jenis_salah	kelas_salah	kata_salah
0	Laki-Laki	1	Vokal	35
1	Laki-Laki	2	Diftong	12
2	Laki-Laki	3	Konsonan	15
3	Perempuan	1	Vokal	29
4	Perempuan	2	Diftong	4
5	Perempuan	3	Konsonan	12

Notes:

*Jenis* : types

*Salah* : false

*Kata* : words

*Laki-laki* : male

*Perempuan* : female

*Vokal* : vocal

*Konsonan* : consonant

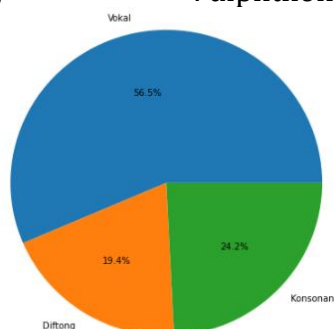
*Diftong* : diphthong

The male of types data errors:

	jk	id_jenis_salah	kelas_salah	kata_salah
0	Laki-Laki	1	Vokal	35
1	Laki-Laki	2	Diftong	12
2	Laki-Laki	3	Konsonan	15

Notes:

*Jenis* : types  
*Salah* : false  
*Kata* : words  
*Laki-laki* : male  
*Perempuan* : female  
*Vokal* : vocal  
*Konsonan* : consonant  
*Diftong* : diphthong



Notes:

*Vokal* : vocal  
*Konsonan* : consonant  
*Diftong* : diphthong

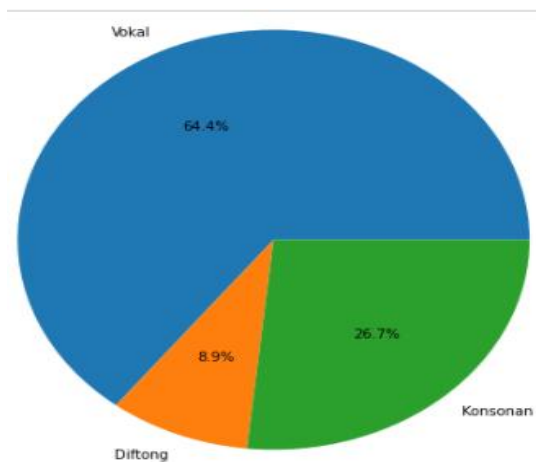
The female of types data errors:



	jk	id_jenis_salah	kelas_salah	kata_salah
3	Perempuan	1	Vokal	29
4	Perempuan	2	Diftong	4
5	Perempuan	3	Konsonan	12

Notes:

*Jenis* : types  
*Salah* : false  
*Kata* : words  
*Perempuan* : female  
*Vokal* : vocal  
*Konsonan* : consonant  
*Diftong* : diphthong



Notes:

*Vokal* : vocal  
*Konsonan* : consonant  
*Diftong* : diphthong

**d. Data Group By Profession**



id_profesi	label_profesi	id_jenis_salah	kelas_salah	kata_salah	
0	1	Guru	1	Vokal	41
1	1	Guru	2	Diftong	10
2	1	Guru	3	Konsonan	14
3	2	Dosen	1	Vokal	23
4	2	Dosen	2	Diftong	6
5	2	Dosen	3	Konsonan	13

Notes:

*Jenis* : types  
*Salah* : false  
*Kelas* : class  
*Kata* : words  
*Profesi* : profession  
*Guru* : teachers  
*Dosen* : lectures  
*Vokal* : vocal  
*Konsonan* : consonant  
*Diftong* : diphthong

### The teachers' data errors:

The all teachers:

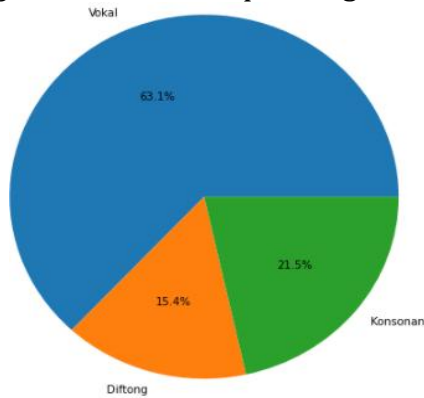
id_profesi	label_profesi	id_jenis_salah	kelas_salah	kata_salah	
0	1	Guru	1	Vokal	41
1	1	Guru	2	Diftong	10
2	1	Guru	3	Konsonan	14

Notes:

*Jenis* : types  
*Salah* : false  
*Kelas* : class  
*Kata* : words  
*Profesi* : profession



*Guru* : teachers  
*Vokal* : vocal  
*Konsonan* : consonant  
*Diftong* : diphthong



Notes:

*Vokal* : vocal  
*Konsonan* : consonant  
*Diftong* : diphthong

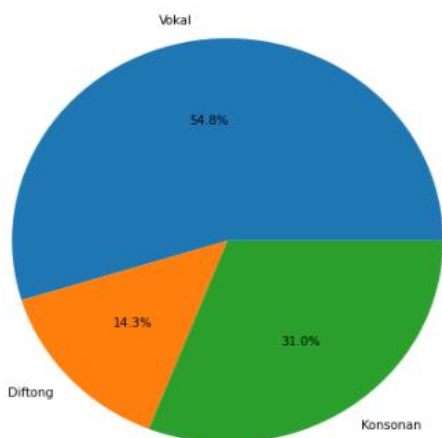
Thea all lectures:

id_profesi	label_profesi	id_jenis_salah	kelas_salah	kata_salah
3	2	Dosen	1	Vokal 23
4	2	Dosen	2	Diftong 6
5	2	Dosen	3	Konsonan 13

Notes:

*Jenis* : types  
*Salah* : false  
*Kelas* : class  
*Kata* : words  
*Profesi* : profession  
*Dosen* : lecturers  
*Vokal* : vocal  
*Konsonan* : consonant  
*Diftong* : diphthong





Notes:

*Vokal* : vocal

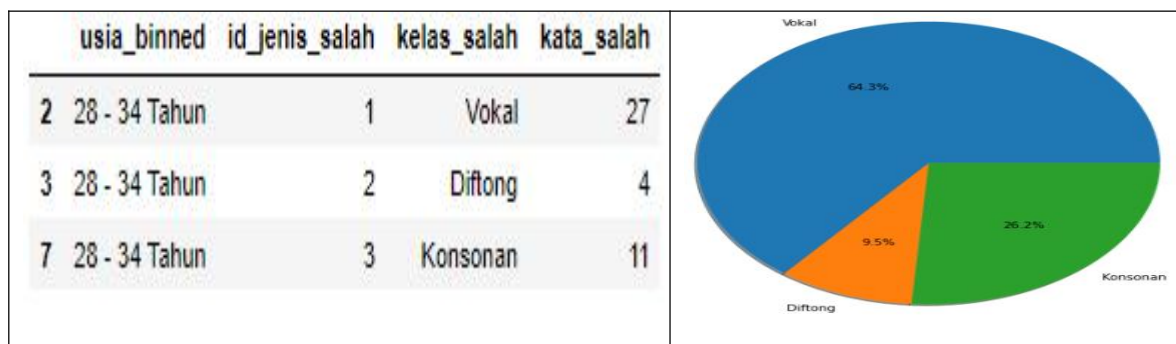
*Konsonan* : consonant

*Diftong* : diphthong

### e. Data Group By Age

usia_binned	id_jenis_salah	kelas_salah	kata_salah
2 28 - 34 Tahun	1	Vokal	27
3 28 - 34 Tahun	2	Diftong	4
7 28 - 34 Tahun	3	Konsonan	11
11 34 - 40 Tahun	1	Vokal	12
12 34 - 40 Tahun	2	Diftong	3
16 34 - 40 Tahun	3	Konsonan	9
20 40 - 46 Tahun	1	Vokal	25
21 40 - 46 Tahun	2	Diftong	9
25 40 - 46 Tahun	3	Konsonan	7

- 28 - 34 Tahun

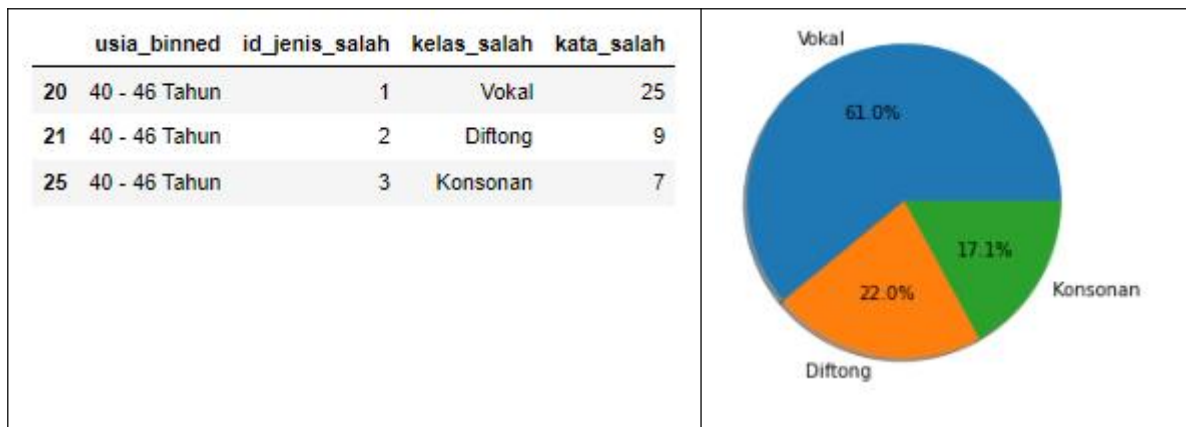




- 34 – 40 Tahun



- 40 – 46 Tahun



### Data Statistics

	id_person	id_jenis_salah	id_profesi	id_tingkat_profesi	usia	kata_salah
count	28.000000	28.000000	28.000000	28.000000	28.000000	28.000000
mean	5.928571	1.964286	1.250000	2.107143	38.500000	3.821429
std	3.640600	0.881167	0.440959	1.196887	6.437736	3.810838
min	1.000000	1.000000	1.000000	1.000000	28.000000	1.000000
25%	3.000000	1.000000	1.000000	1.000000	30.000000	2.000000
50%	5.000000	2.000000	1.000000	2.000000	40.000000	3.000000
75%	9.250000	3.000000	1.250000	2.500000	44.000000	4.000000
max	12.000000	3.000000	2.000000	4.000000	46.000000	19.000000



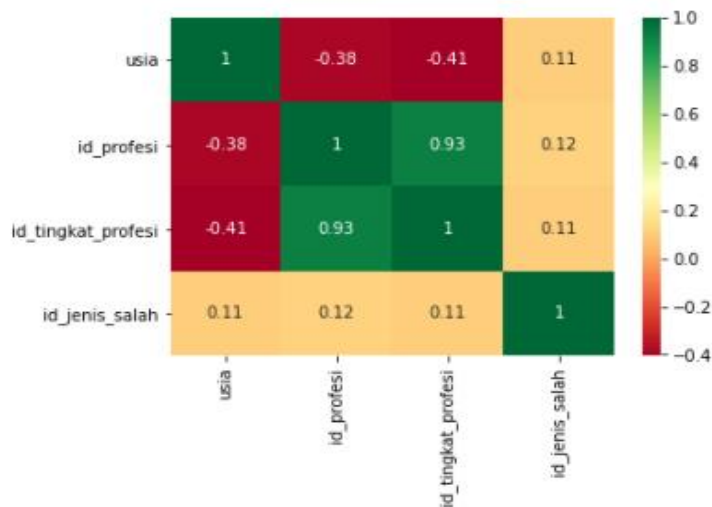
Chi2 score attribute to the types of word errors:

	Atribut	Chi2 Score
0	id_profesi	0.068052
1	id_tingkat_profesi	0.351618
2	jk	0.048804
3	usia	0.363569

Correlation Matrix:

	usia	id_profesi	id_tingkat_profesi	id_jenis_salah
usia	1.000000	-0.384882	-0.406169	0.107728
id_profesi	-0.384882	1.000000	0.929825	0.119149
id_tingkat_profesi	-0.406169	0.929825	1.000000	0.109115
id_jenis_salah	0.107728	0.119149	0.109115	1.000000

Confusion Matrix:



Notes:

- Jenis* : types
- Tingkat* : level
- Usia* : age
- Salah* : false
- Profesi* : profession



*Kelas* : class  
*Kata* : words  
*Tahun* : years  
*Vokal* : vocal  
*Konsonan* : consonant  
*Diftong* : diphthong

## CONCLUSION

From the data analysis above, it can conclude that the mapping and statistical information obtained can be used as a basis in developing programs that aim to improve the quality of the use of standard Indonesian language in the form of FGD, socialization and counseling as well as in the form of technology products. The results in this research indicate standard pronunciation errors that occur in various circles including the presenters and lecturers. After that, the results of data analysis are described in the form of tables and diagrams so that it is seen the percentage of errors that occur.

## REFERENCES

- Anusuya, M., & Katti, S. (2009). Speech Recognition by Machine: A Review. International Journal of Computer Science and Information Security.
- Badudu, J.S. (1993). Cakrawala Bahasa Indonesia, Jakarta: Gramedia.
- Benkerzaz, S., Dennai, A., Elmir, Y. (2019). A Study on Automatic Speech Recognition. Department of Exact Sciences, University of TAHRI Mohamed Smart Grid & Renewable Energies Laboratory, Computer Science & Sciences Didactic Team Bechar, Algeria.
- Brown, P., desouza, P., Mercer, R., Pietra, V., & Lai, J. (1992). Class-Based n-gram Models of Natural Language. New York: IBM T. J. Watson Research Center.
- Das, P & Prasad, V. (2015). Voice Recognition System: Speech-To-Text. Journal of Applied and Fundamental Sciences. Department of Computer Science & Engineering and Information Technology, School of Technology, Assam Don Bosco University, Assam, India
- Erwina, Emmy (2012). Kajian Sebutan Baku Bahasa Indonesia. Singapore. Singapore International Press.
- Fahrudin, T., Buliali, J., & Fatichah, C. (2019). Ina-BWR: Indonesian bigram word rule for multi-label student complaints. Egyptian Informatics Journal.



Jurafsky, D., & Martin, J. (2018). *Speech and Language Processing*. Stanford University.

Mawardi, Viny Christanti. (2020). *Jurnal Fast and Accurate Spelling Correction Using Trie and Bigram*. SCOPUS. Universitas Ahmad Dahlan dan IAES.

Mustikarini, W., Hidayat, R., & Bejo, A. (2019). *Real-Time Indonesian Language Speech Recognition with MFCC Algorithms and Python-Based SVM*. IJITEE, Vol. 3, No. 2.

Nisa, Khairun. (2018). *Analisis Kesalahan Berbahasa Pada Berita Dalam Media Surat Kabar Sinar Indonesia Baru*. *Jurnal Didaktika FKIP UM Palembang*.

Samanta, P., & Chaudhuri, B. (2013). *A simple real-word error detection and correction using local word bigram and trigram*. *Proceedings of the Twenty-Fifth Conference on Computational Linguistics and Speech Processing (ROCLING)*.

Sallyanti, Dra. (2003). *Analisis Bahasa Baku Dan Non Baku Dalam Bahasa Indonesia*. USU digita library. Fakultas Sastra Jurusan Bahasa Dan Sastra Indonesia Universitas Sumatera Utara.

Sukarta, N., Suparwa, N., Putrayasa, G.N.K., & Teguh, W. (2015). *Bahasa Indonesia Akademik untuk Perguruan Tinggi*. Udaya University Press.

Trivedi, A., Pant, N., Shah, P., Sonik, S., & Agrawal, S. (2018). *Speech to text and text to speech recognition systems-A review*. IOSR Journal of Computer Engineering. e-ISSN: 2278-0661,p-ISSN: 2278-8727.

Wahyuni, T, L. (2010). *Analisis Pemakaian Bahasa Indonesia Penyiar Radio Pop FM Sragen Dalam Acara School Hoppers*. Thesis Pendidikan Bahasa, Sastra Indonesia Dan Daerah Fakultas Keguruan Dan Ilmu Pendidikan Universitas Muhammadiyah Surakarta.