



Final Defense

on

Generating Opinion Lexicon for Nepali using Microblogs

Presented by

Sawan Vaidya

(15526)

Supervisor: Asst. Prof. Bal Krishna Bal



Background

❖ SENTIMENT ANALYSIS (OPINION MINING)

Definition from Google:

noun

the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral.

"companies have key lessons to learn about harnessing the power of social media and sentiment analysis"



Background

❖ MICRO-BLOGS

Definition from Google:

noun

a social media site to which a user makes short, frequent posts.

"she announced her retirement on her microblog, bidding farewell to the sport"

Examples: Twitter, Youtube Comments, Ecommerce Reviews, Article Comments etc



Background

❖ LEXICON – 2 WORD LEXICON

word	positive	negative
राम्रो	0.9	0.1
नराम्रो	0.1	0.9

=

word	Score
राम्रो	0.8
नराम्रो	-0.8

$\text{score} = \text{positive} - \text{negative}$

$1 = \text{positive} + \text{negative}$

❖ LEXICON – EMOJI SENTIMENT DATA[4]

Emoticon	Positive	Negative	Neutral	Emoticon	Positive	Negative	Neutral
😊	0.08	0.79	0.13	😬	0.68	0.05	0.27
😐	0.08	0.79	0.13	💡	0.60	0.04	0.36
😓	0.08	0.79	0.13	🏠	0.66	0.11	0.23
😞	0.06	0.75	0.19	🏆	0.62	0.09	0.29
😔	0.10	0.78	0.12	💞	0.61	0.09	0.30
😡	0.13	0.77	0.09	♣️	0.63	0.11	0.26
😐	0.08	0.72	0.20	👉	0.62	0.11	0.27
😓	0.15	0.77	0.08	❤️	0.63	0.13	0.24
👊	0.13	0.73	0.15	🏃	0.62	0.12	0.25
👊	0.12	0.72	0.16	💞	0.59	0.12	0.29



Background

❖ LEXICON – 32 WORD MANUAL

word	positive	negative	neutral	word	positive	negative	neutral	word	positive	negative	neutral
राम्रो	0.9	0.05	0.05	शान्ति	0.9	0.05	0.05	जबरजस्ती	0.05	0.9	0.05
स्वच्छ	0.9	0.05	0.05	शुद्ध	0.9	0.05	0.05	हस्तक्षेप	0.05	0.9	0.05
चम्किलो	0.9	0.05	0.05	स्वस्थ	0.9	0.05	0.05	नकारात्मक	0.05	0.9	0.05
सभ्य	0.9	0.05	0.05	नियमित	0.9	0.05	0.05	फटाहा	0.05	0.9	0.05
सफा	0.9	0.05	0.05	व्यवस्थित	0.9	0.05	0.05	लुटेरा	0.05	0.9	0.05
सक्षम	0.9	0.05	0.05	नराम्रो	0.05	0.9	0.05	चोर	0.05	0.9	0.05
शान्त	0.9	0.05	0.05	भ्रष्ट	0.05	0.9	0.05	वाक्क	0.05	0.9	0.05
पौष्टिक	0.9	0.05	0.05	सुस्त	0.05	0.9	0.05	आतंक	0.05	0.9	0.05
ताजा	0.9	0.05	0.05	हचुवा	0.05	0.9	0.05	तोडफोड	0.05	0.9	0.05
निष्पक्ष	0.9	0.05	0.05	अनियमित	0.05	0.9	0.05	अपहरण	0.05	0.9	0.05
विश्वसनीय	0.9	0.05	0.05	झुर	0.05	0.9	0.05				



Problem Statement

❖ STATEMENT

“A basic ingredient of **Sentiment Analysis** that uses Bag-of-words approach is an **Opinion Lexicon** containing words/phrases along with their positivity and negativity ratings. Many methods have been suggested for building such a lexicon. In this thesis a method **utilizing micro-blogs** (user comments from Facebook pages) is suggested. While, this lexicon will be **domain specific**, it will be a **first in the Nepalese context** and can lay the foundation for development of Sentiment Analysis resources. Major **challenges** are that data generated by users on the Internet is **subject to informal language**. Also **misspellings, abbreviations, emoticons** and **multi-language** content pose additional hurdles. The benefits are that **micro-blogs are richer in subjective content** compared to text found in formally written sources. The thesis will investigate the use existing lexicons (such as emoticon lexicon) and manually generated lexicons as seed lexicon for enriching and expanding them..” – Text from Report



Objectives

- Develop a method of generating opinion lexicon using text available in microblogs
- Refine the strategy for filtering out mistakes and useless terms from the microblog lexicon
- Generate an opinion lexicon for Nepali, specify its pros and cons.



Literature Review

❖ TWO PREVALENT METHODS

- Dictionary Based

- Uses word relations in dictionaries (like WordNet [3])
- Eg. SentiWordNet [1]

- Corpus Based

- Analyses occurrences of words in documents
- Mostly uses Pointwise Mutual Information between words.
Eg: [5]



Literature Review

❖ PRIMARY REFERENCE

Paper: Detecting Sentiment in Nepali Texts [2]

- Translation of SentiWordNet words to Nepali (Bhavanakos)
- Sentiment Clue List
- Nepali Sentiment Corpus

❖ LIMITATIONS

- Translated from English Lexicon
- Not representative of language used in Microblogs
- Absence of Polarity Strengths



Literature Review

❖ OTHER INSPIRATIONAL PAPERS

Paper: Sentiment of Emojis [4]

- Polarity strengths for various Emojis

Paper: Sentiment lexicon generation for an under-resourced language [5]

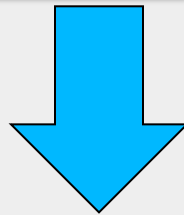
- Selects top patterns in text
- Uses Pointwise Mutual Information between words

Methodology

❖ EXAMPLE (TEXT CLASSIFICATION)

Word	Pos (p)	Neg (n)
राम्रो	0.9	0.1
नराम्रो	0.1	0.9

Primary Lexicon



Use

Classification

Sentences

Text	Average score(p)	verdict
म राम्रो र सफा	0.8	Positive
म नराम्रो र फोहर	-0.8	Negative
म सफा र शुद्ध	0	Unknown
म फोहर र गनाउने	0	Unknown
म राम्रो र नराम्रो	0	Unknown

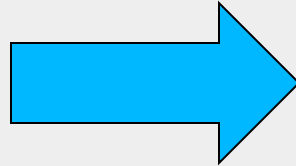
Methodology

❖ EXAMPLE (COUNTING)

Text	verdict
म राम्रो र सफा	Pos
म नराम्रो र फोहर	Neg

Classification

Counts



Count Words

	<u>Counts</u>	
Word	Pos	Neg
राम्रो	1	0
सफा	1	0
म	1	1
र	1	1
नराम्रो	0	1
फोहर	0	1

❖ CALCULATE STRENGTH

Word (w)	Pos	Neg	P (w Pos)	P (w Neg)	Strength (pos)	Strength (neg)	Score
राम्रो	1	0	0.25	0	1.00	0.00	1
<u>सफा</u>	1	0	0.25	<u>0</u>	<u>1.00</u>	<u>0.00</u>	<u>1</u>
म	1	1	0.25	0.25	0.50	0.50	0
र	1	1	0.25	0.25	0.50	0.50	0
नराम्रो	0	1	0	0.25	0.00	1.00	-1
<u>फोहर</u>	0	1	0	<u>0.25</u>	<u>0.00</u>	<u>1.00</u>	<u>-1</u>
Total	4	4					

New Words found

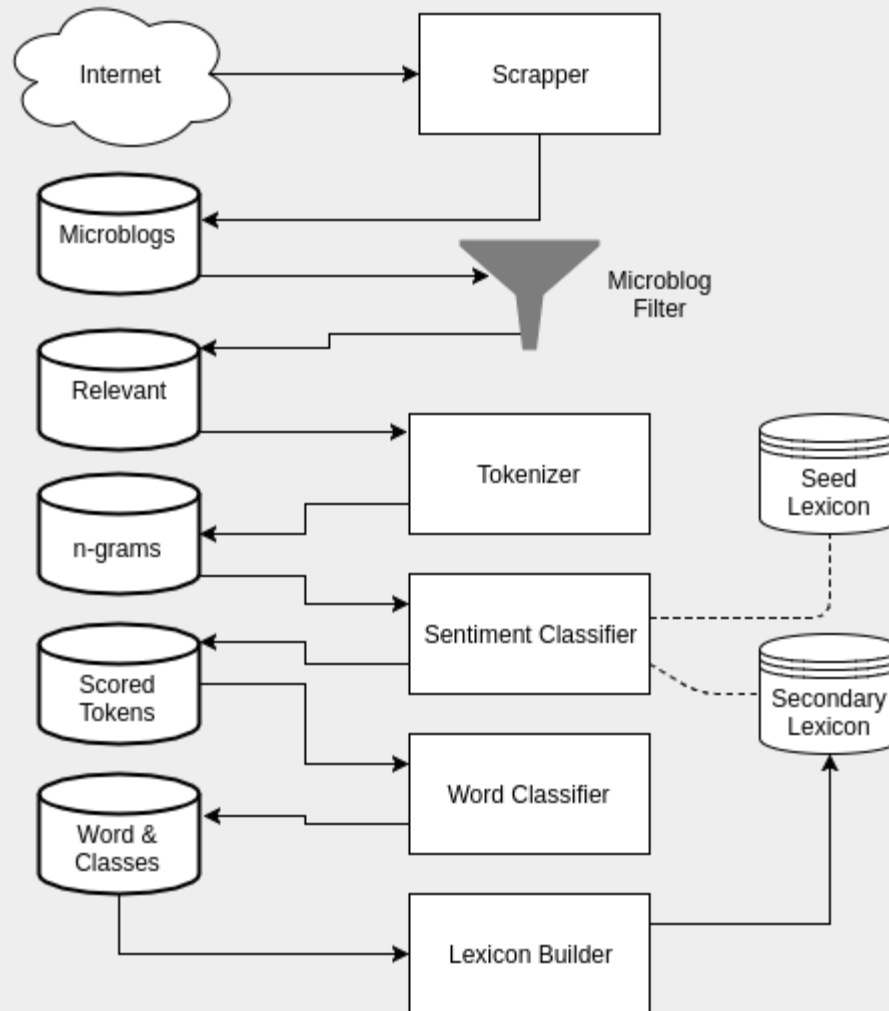


Methodology

❖ ASSUMPTIONS

- Bag of Words
- Neutral words occur in both positive and negative contexts
- Positive words occur more with other positive words
- Negative words occur more with other negative words

Methodology





Methodology

❖ DATA COLLECTION (FACEBOOK PAGES)

PageID	Comments	PageID	Comments
MaoistCenterNp	2901	bbcsajhasawal	34320
BibeksheelSajha	7590	NepalPolicePHQ	38768
yourMadanKrishnaShrestha	8433	dc.nepal	42433
nepalicongresshq	9864	DrBaburamBhattarai	51384
communistpartynepal	11986	rekhatapa.net	55116
hari.bamsha.acharya	14286	PriyankaKarkiofficial	60660
setopati	15162	BBCnewsNepali	64062
nepal8thwonder	23158	nagariknews	71181
onlinekhabarnews	30267	eKantipur	73286



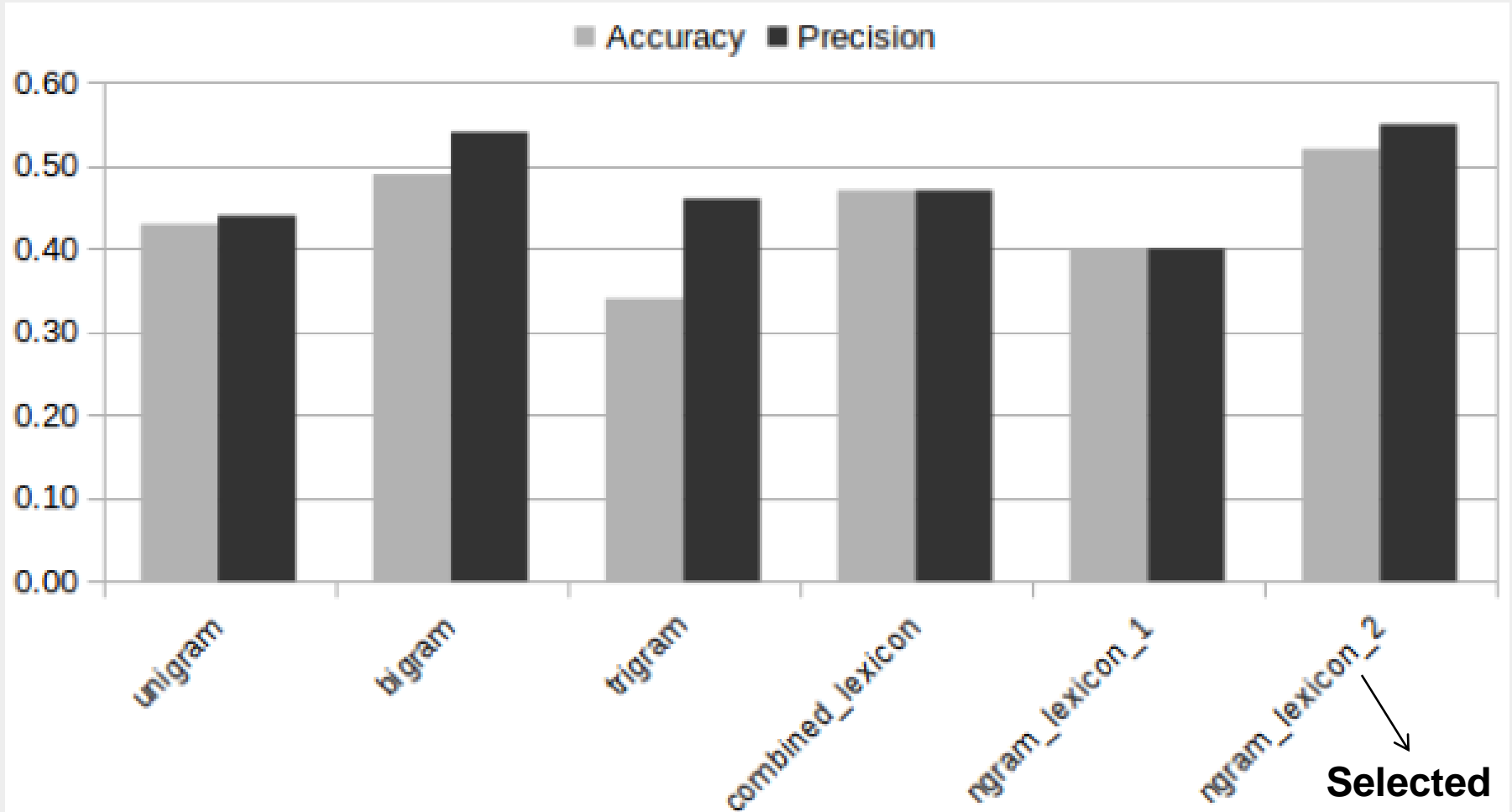
Methodology

❖ CHARACTER SELECTION

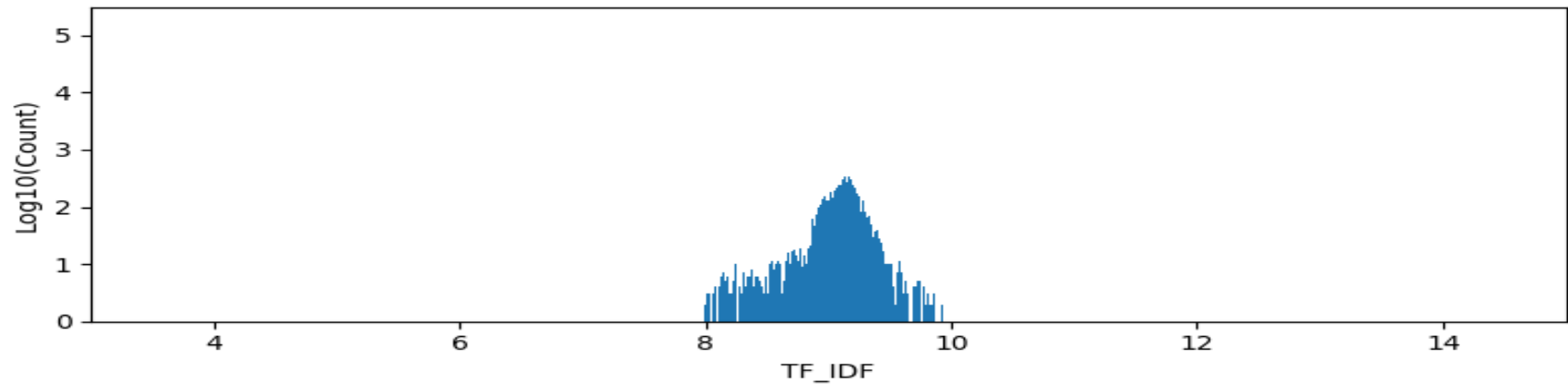
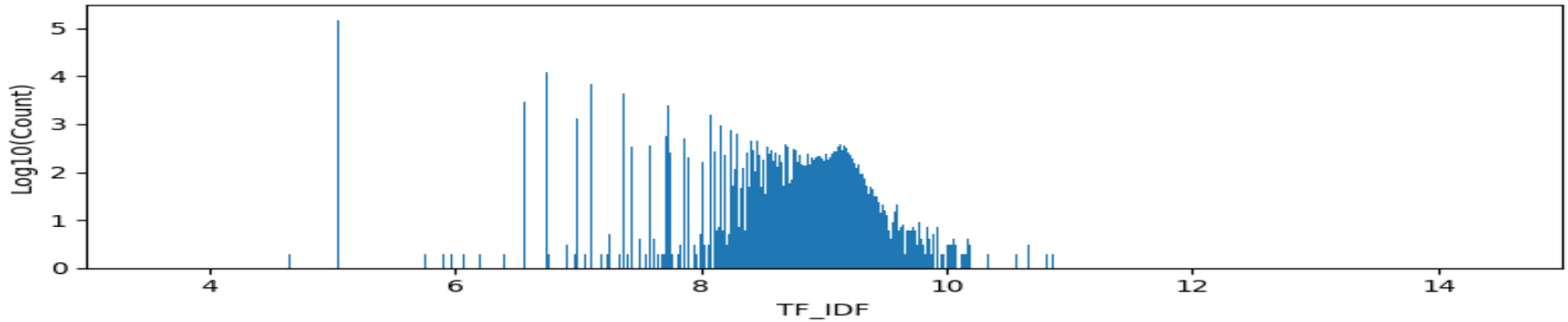
Latin	Devanagari	Emoticons	Count
No	No	No	16694
No	No	Yes	890
No	Yes	No	106506
No	Yes	Yes	3180
Yes	No	No	448296
Yes	No	Yes	11317
Yes	Yes	No	26227
Yes	Yes	Yes	1747
No	Yes	<u>Don't Care</u>	109686

Dataset that was used in bold
1059 Test data was subtracted from it

❖ TOKENIZATION METHODS



❖ PRE FILTER : $8 < \text{TF-IDF} < 10$ AND $\text{DOC-COUNT} > 32$

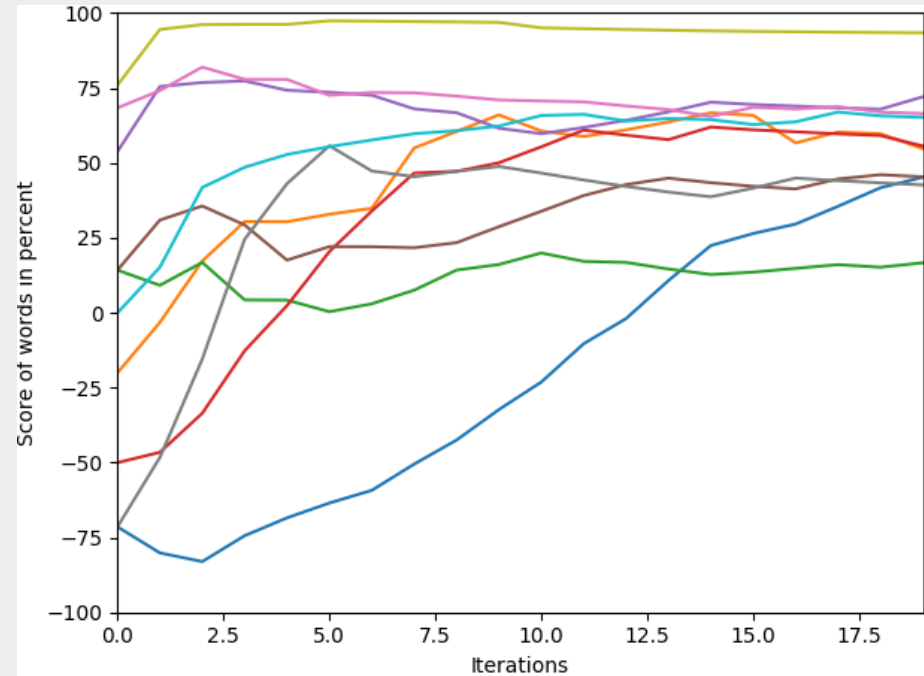
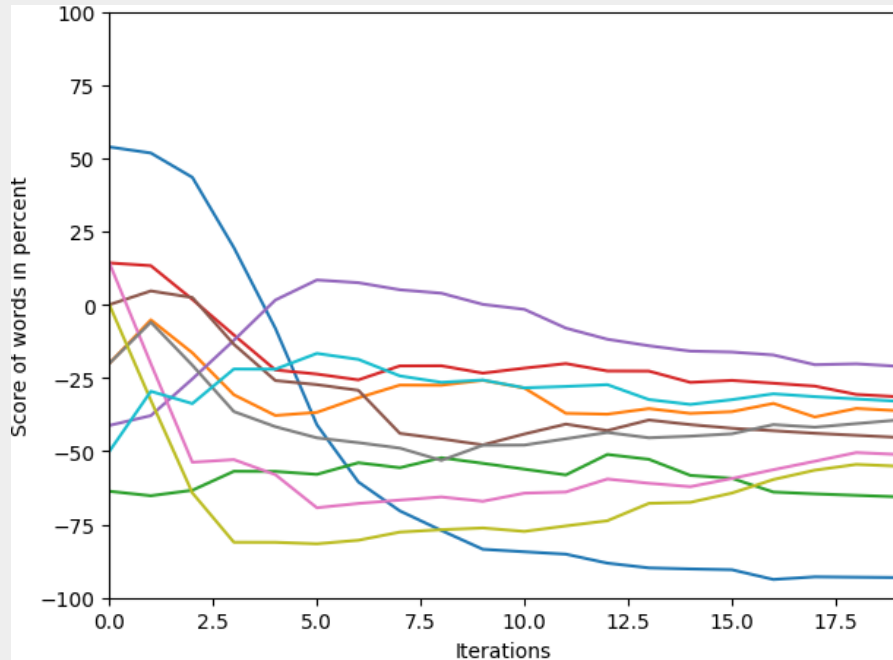




Methodology



ITERATION EFFECT





❖ BUILD FILTERS : FILTERS AT LEXICON BUILDER STAGE

Skewness	$Skewness = \frac{3(mean - median)}{stdev}$
Range	$highest\ score - lowest\ score$
Overall Score	The final score obtained by an ngram
Count	The number of observations for forming a score



Experiments

Experiment	Seed Lexicon
Experiment 1	Emoticon
Experiment 2	32 word
Experiment 3	2 word



Experiments

❖ INITIAL PARAMETERS

Parameter	Value	Parameter	Value
<u>Devanagari</u>	Mandatory	<u>TF_IDF min</u>	8
<u>Latin</u>	Disallow	<u>TF_IDF max</u>	10
<u>Test Data</u>	Disallow	<u>Doc count min</u>	32
<u>Emoticons</u>	Optional	<u>Iterations</u>	20
<u>TestData</u>	1059	<u>Ngrams</u>	Ngram Method 2
<u>Microblog Count</u>	108566	<u>Build Parameters</u>	See Table 4.2

<u>Range</u>	$\text{range} \geq 0.05$
<u>Count</u>	$\text{count} \geq \text{Mean}/4$
<u>Score</u>	$ \text{score} \geq 0.5$
<u>Skewness</u>	$ \text{skewness} \geq \text{Mean}/8$



Evaluation

❖ TEST DATA : SAMPLE

Text	Sentiment
म मंगल ग्रह बाट सुन्दै छु	0
प्रधानमंत्री को भारत यात्रा ले नेपाल को फाइदा होइन नोकसान मात्र हुने छ किनकी देउवा भारत को शिष्य हो। कुनै पनि संधि समझेर उता नगरोस भारत संग	-1
हाम्रो देशको बाढी पहिरो कस्ले हेरिदिने कस्ले बुझिदिने ??	-1
धन्य हो हाम्रो नेपाली सेना सबै थाउ मा उहाँ हरू सहयोग ले नै देस बचि राख मेरो सलाम छ	1



Evaluation

❖ TEST DATA: COMPOSITION

Total Unique Tokens	10277
Total 1 grams matching TF-IDF criteria	4546
Total 2 grams matching TF-IDF criteria	2195
Total 3 grams matching TF-IDF criteria	246
Positive Comments	353
Negative Comments	353
Unknown and Neutral Comments	353



Evaluation

❖ CONFUSION MATRIX : RANDOM CLASSIFICATION

Class	Precision	Recall	F-score
A	0.34	0.33	0.33
B	0.36	0.35	0.35
C	0.3	0.32	0.31
Overall Accuracy	0.33	Mean Precision	0.35

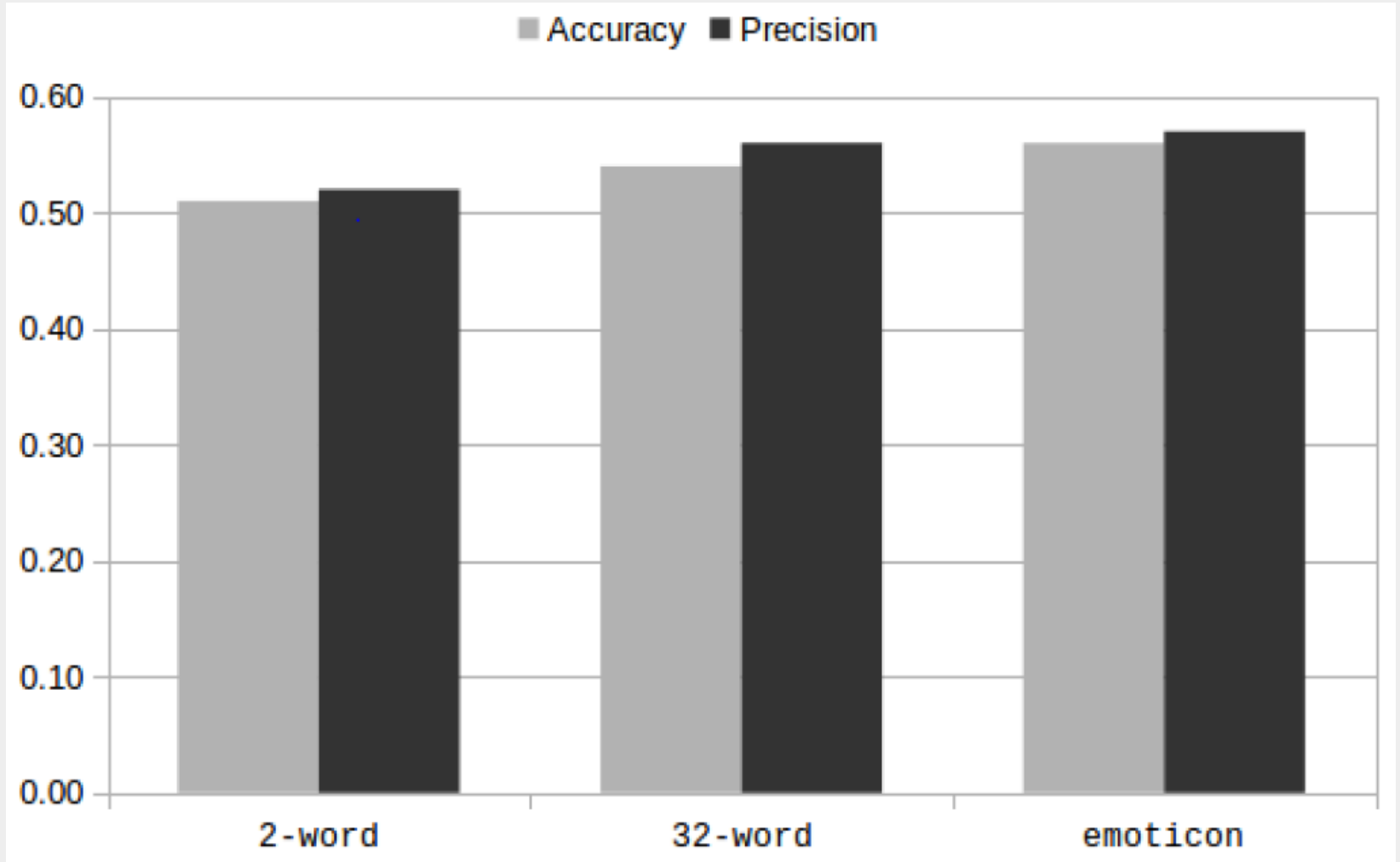


❖ ACCURACY AND PRECISION

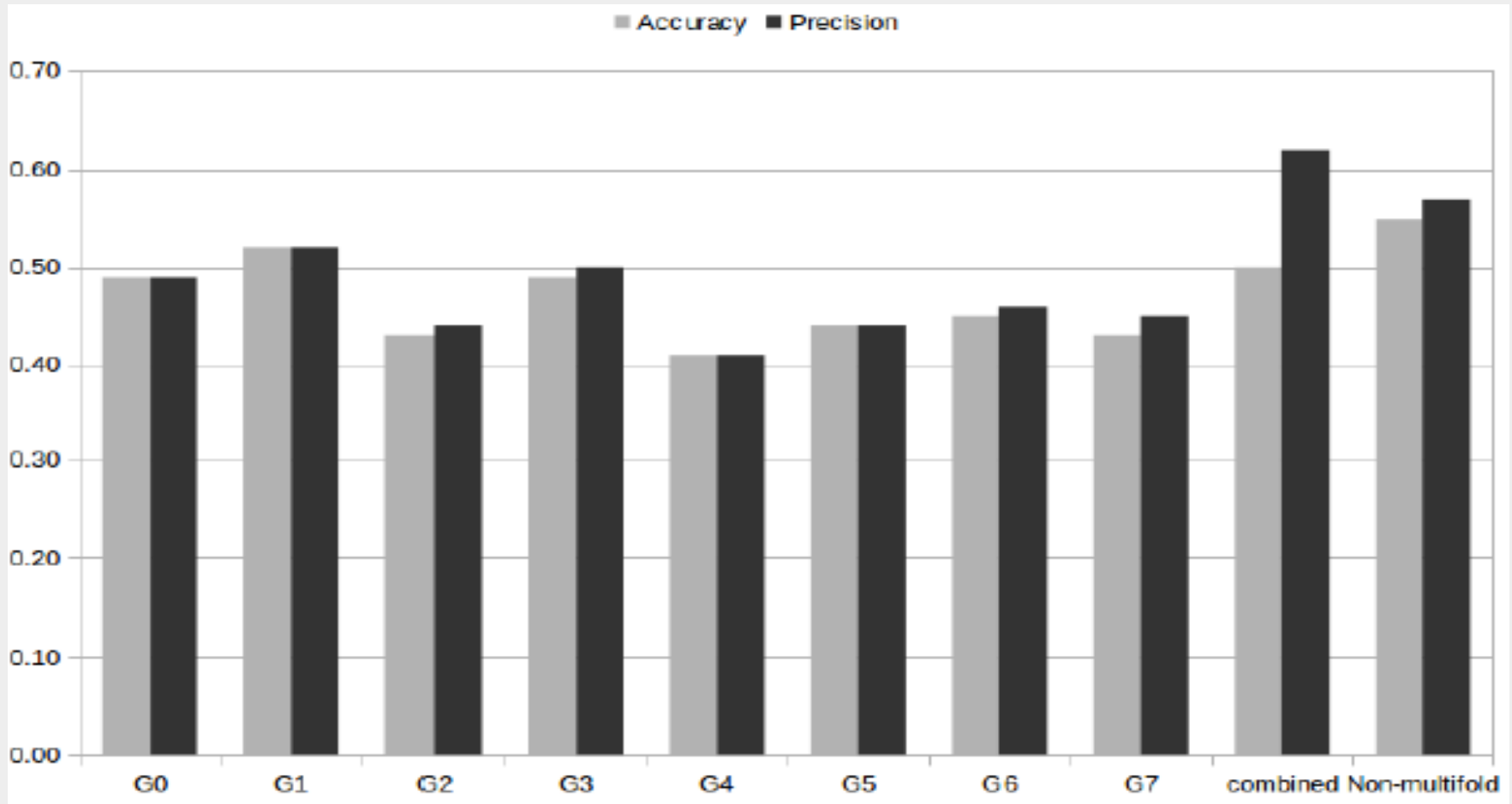
$$\text{Overall Accuracy} = \frac{\text{Correct Classifications}}{\text{Total Classifications}}$$

$$\begin{aligned} \text{Mean Precision} \\ = \sqrt{\text{Precision (Pos)} * \text{Precision (Neg)}} \end{aligned}$$

❖ EXPERIMENTS COMPARED



❖ MULTIFOLD ANALYSIS (EMOTICON SEED)





Trigrams (2-word Seed)

के.पी.ओली केही हुनेवाला छैन नेपाल जय नेपाली
लागु जे पनियद तिपाई लाई
मुटुमा बस्न सफल दानु परछ भन्ने कुनै पन हालतमा
कुनै औचित्य छैन हो कः कः
मन छुने नेपाली धारण गर्ने शक्ति
खै के खै लाईक गरि भ्रष्टाचार
छ कः छैन नेपाली जनता र
प्रचण्ड र बाबुराम दानि एक मात्र
के छ र माया गर्ने र बाहेक अरु केही
जे पन गिरन
पन कानै चरिको सबै भन्दा ठुलो हो कः कः
धैर्य धारण गर्नु आए पन का
बुझनु जुरी छ बिकास र समृद्धिको
रगत र पसनि जय नयाँ शक्ति देश र जनताका
के का लागि भने भने एक पटक कः कः
मन परछ भने नेकपा एमाले जनिदावाद
प्रत्यक्ष न रिवाचति कार्यकारी
बनाउने हो भने
के नै गर्नु राजा महेन्द्र ले
अरु केही होइन हुन्छ भने एक नै कः पा अरु कः
हुदैन तब सम्म नै छ र कः कः
एक पटक यो कानमा तेल हालेर
कांग्रेस र एमाले भनेको यही हो बुझनु जुरी छ
मन मुटुमा बस्न हाम्रो देश नेपालमा के हेरेर बसेको
कः कः के यद छ भने त के कुरा
के को लागि पन हुन सक्छ नेपाल को कानुन
चरि शान्तिको कामना यो पेज लाईक
कः पा एमाले नेकपा एमाले जनिदावाद
आत्माको चरि शान्तिको गर्नु पर यो
हो भने हामी हुनु पर्ने हो जय नेपाल जय
के हुन सक्छ जोगी आए पन
स्वास्थ्य लाभको कामना भनेको यहि हो यो साझा सवाल नेपाल प्रहरी ले

ले बधाई छ रेखा थापा जी छ नेपाल मा
कुरा गर्नु भयो नेपाल पुलसि लाई को कुरा हो
हजिको साझा सवाल देश र जनताको
डा बाबुराम भट्टलाई गर्नु भएको छ भए राम्रो हुन्छ
हार्दकि मंगलमय शुभ को कुरा सुनदा हार्दकि बधाई छ हो कः जस्तो
राम्रो हुने थियो बधाई तथा सफल
सतलि सरापेको देश धेरै धेरै धन्यवाद हार्दकि बधाई तथा लाई धेरै धेरै
हुने थियो कः रेखा जी को को हार्दकि मंगलमय
बधाई तथा शुभकामना र जनताको लागि नेपाल प्रहरी को
हार्दकि मंगलमय शुभकामना
आजको साझा सवाल कुरा हो तर छ नेपाल प्रहरी
तथा सफल कार्यकालको
बिबिसि राम्रो काम गर्नु राम्रो काम को
राम्रो हुन्छ यो कः धेरै राम्रो काम
होला जस्तो छ कुरा हो र धेरै धेरै बधाई
बी बी सी मेरो बचिार मा होला जस्तो लाग्छ
लाई बधाई छ सलाम छ नेपाल हाम्रो नेपाल मा
आखामा छारो हालेर को साथ मा जन्म दानि को धेरै धेरै बधाई
रेखा थापा को साझा सवाल मा साझा सवाल ले
राम्रो काम गर्नु राम्रो काम हो राम्रो काम गरेर
काम गरेर देखाउनु सफल कार्यकालको शुभकामना
बाबुराम भट्टलाई लाई
धेरै राम्रो लाग्यो बधाई छ कमरेड को खाचो छ धेरै बधाई छ
नै धर्म हो बिबिसी धेरै राम्रो छ धेरै धेरै धन्यवाद
साझा सवाल को मेरो तरफ बाट धेरै धेरै धन्यवाद
को साझा सवाल रेखा मैसाप जी बाबुराम जति
राम्रो काम गर्नु जय जय साझा सवाल कार्यकर्म
धेरै धेरै शुभकामना काम गरेर देखाउ शुभकामना व्यक्त गर्द
राम्रो हो तर काम गर्नु लाई राम्रो काम गरे
राम्रो कुरा हो नेपाल प्रहरी लाई
काम गर्नु परछ यो साझा सवाल नेपाल प्रहरी ले



Trigrams (32-word Seed)

अब.नेपाल.मा लाई.के.थाहा नेपाल.आमा.को
पर.यो.न कि.गर्ने.त का.नेता.हरु
सबैलाई.चेतना.भया
कांग्रेस.र.एमाले परछ.अनि.मात्र
के.यही.हो राजा.महेन्द्र.ले
अनु.के. भनेको.यही.हो के.नै.गर्न
प्रचण्ड.र.बाबुराम नेता.हरु.ले
के.पी.ओली त.होला.न
आखामा.छारो.हालेर
भनेको.यही.हो नेपाल.मा.पनि
कानमा.तेल.हालेर
क.पा.एमाले के.का.लागि
के.हुन.सक्छ ने.क.पा
दुनि.परछ.भन्ने नै.छ.र त.के.कुरा
सबै.भन्दा.धेरै
नेपाल.को.नेता गर्नु.पर.यो
हो.कि.क्या लागि.जे.पनि हाम्रो.देश.नेपालमा
जे.पनि.गर्न त.हो.न निता.हरु.को
र.जनता.लाई
नेपाल.को.कानून स्वार्थ.को.लागि
के.हेरेर.बसेको को.हात.मा
रगत.र.पसनि के.कुरा.गर्नु
नेपाल.का.नेता केही.फरक.पर्दैन

नेपाल.प्रहरी.को डा.बाबुराम.भट्टराई काम.गरेर.देखाउ
साझा.सवाल.कार्यक्रम
हार्दकि.बधाई.छ लाई.धेरै.धेरै को.कुरा.हो
जन्म.दनि.को चरि.शान्तिको.कामना
रेखा.धापा.जी र.जनताको.लागी
धेरै.धेरै.धन्यवाद काम.गरेर.देखाउनु रेखा.धापा.को
हुन्छ.भने.एक पेज.लाई.क.गर्न
धारण.गर्ने.शक्ति राम्रो.हो.तर एक.पटक.यो
माया.गर्ने.र आत्माको.चरि.शान्तिको
लाई.क.गर्न.भ्रष्टाचार साझा.सवाल.को राम्रो.काम.गर्न
जय.नयाँ.शक्ति राम्रो.हुन्छ.यो.कि
धेरै.राम्रो.छ साझा.सवाल.मा गर्नु.भएको.छ
हमिल.पहाड.तराई जय.जय.जय काम.गर्ने.लाई
रेखा.मैसाप.जी राम्रो.काम.हो
राम्रो.हुने.थियो कुनै.न.कुनै जस्तो.लाग्न.छ.लाई धेरै.बधाई.छ
नेपाल.प्रहरी.ले राम्रो.काम.को धेरै.धेरै.धन्यवाद
बधाई.छ.कमरेड मेरो.तरफ.बाट धेरै.धेरै.शुभकामना
को.हार्दकि.मंगलमय हामी.सबै.नेपाली
दनि.एक.मात्र बाबुराम.भट्टराई.जयु न.बि.स
हो.भने.हामी स्वास्थ.लाभको.कामना यो.साझा.सवाल
आजको.साझा.सवाल को.साथ.मा साझा.सवाल.ले
प्रत्यक्ष.न.रिवाचति.कार्यकारी को.हो.र नयाँ.शक्ति.नेपाल सलाम.छ.नेपाल
हार्दकि.मंगलमय.शुभकामना
नेपाल.प्रहरी.लाई राम्रो.कुरा.हो राम्रो.काम.गर्ने धेरै.राम्रो.लाग्यो
गरे.कसो.होला ब.बि.सी धेरै.राम्रो.काम
राम्रो.काम.गरेर यद.छ.भने.छ.नेपाल.प्रहरी हार्दकि.बधाई.तथा
हुने.थियो.कि नेपाल.पुलिस.लाई बधाई.तथा.सफल
हजिको.साझा.सवाल धैर्य.धारण.गर्नु
बाबुराम.भट्टराई.को बी.बी.सी
हार्दकि.मंगलमय.शुभ
कुल.गर्नु.भयो को.साझा.सवाल काम.गर्नु.परछ
तथा.सफल.कार्यकालको यद.तिपाई.लाई
भने.एक.पटक लाई.बधाई.छ ल.बधाई.छ ब.बि.सी
धेरै.धेरै.बधाई रेखा.जी को भए.राम्रो.हुन्छ.यो
कुरा.हो.र शुभकामना.व्यक्ति.गर्नु सफल.कार्यकालको.शुभकामना



Trigrams (Emoticon Seed)

हुने.वाला.छैन के.के.न नेपाल.सरकार.को रगत.र.पसनि नै.हो.भने
को.हात.मा कानमा.तेल.हालेर अरु.के. के.हुन.सक्छ त.होला.न
नेता.हरु.को नेता.हरु.लाई बाहेक.अरु.केही लाई.के.थाहा का.नेता.हरु
के.खै.के नेपाल.सरकार.ले हुदैन.तब.सम्म पन.किनै.चरिको क.पा.एमाले
सबैलाई.चेतना.भया के.को.लागी आखामा.छारो.हालेर ठकै.हो.तर
राजा.महेन्द्र.ले तर.के.गर्ने ने.क.पा हामिमत.छ.भने यस्तै.हो.भने
भनेको.यह.होआए.पन.कि.का के.हो.भनेर यद.तिपाई.लाई हुन्छ.भने.एक
के.यह.हो भनेको.यही.हो कुरा.के.भने छ.नेपाल.प्रहरी सफल.कार्यकालको.शुभकामना भएको.छ.र
हामी.नेपाली.हो नै.छ.र को.काम.छैन को.लागिमात्र हो.भने.त लाई.बधाई.छ साझा.सवाल.कार्यक्रम
को.नाम.मा त.ठकै.हो को.हार्दकि.मंगलमय धेरै.राम्रो.काम बाबुराम.भट्टराई.को
हाम्रो.नेपाल.मा अब.नेपाल.मा सलाम.छ.नेपाल धेरै.धेरै.धन्यवाद बधाई.तथा.सफल को.साझा.सवाल
के.गर्ने.त देशका.नद.नाला मन.छुने.नेपाली कुरा.गर्नु.भयो स्वास्थ्य.लाभको.कामना धेरै.राम्रो.छ
हाम्रो.देश.नेपालमा के.हेरेर.बसेको बी.बी.सी ब.बि.सी नेपाल.प्रहरी.लाई
के.नै.गर्न खै.के.खै हार्दकि.बधाई.छ बकिस.र.समृद्धिको हार्दकि.मंगलमय.शुभकामना हजिको.साझा.सवाल
के.हो.के त.हो.न रेखा.थापा.जी नेकपा.एमाले.जनिदाबाद हार्दकि.मंगलमय.शुभ
अरु.केही.होइन नेता.हरु.ले आजको.साझा.सवाल धेरै.धेरै.बधाई.छ.कमरेड
को.हो.र जोगी.आए.पनि राम्रो.काम.को हो.भने.हामी राम्रो.कुरा.हो
केही.फरक.पर्दैन नेपाल.का.नेता को.लागि.हो हार्दकि.बधाई.तथा.चरि.शान्तिको.कामना



Conclusion

PROS

- Suitable for micro-blog domain
- Small number of pre-requisites
- Captures colloquial language features

CONS

- Micro-blog domain dependent
- Unable to distinguish POS
- Difficult to test



Conclusion

❖ ACHIEVEMENTS

- Micro-blog domain specific lexicon
- Strategies for refining were studied
- Around 4000 terms could be produced from a simple seed lexicon

❖ FURTHER WORK

- Increasing data size can improve accuracy and precision
- Build a domain independent lexicon using WordNet or digital dictionary



References

- 1 Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In LREC, volume 10, pages 2200--2204, 2010. 4
- 2 Chandan Prasad Gupta and Bal Krishna Bal. Detecting sentiment in nepali texts: A bootstrap approach for sentiment analysis of texts in the nepali language. In Cognitive Computing and Information Processing (CCIP), 2015 International Conference on, pages 1--4. IEEE, 2015. 6, 26
- 3 George A Miller. Wordnet: a lexical database for english. Communications of the ACM, 38(11):39--41, 1995. 4
- 4 Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. Sentiment of emojis. PloS one, 10(12):e0144296, 2015. 7
- 5 Clara Vania, Moh. Ibrahim, and Mirna Adriani. Sentiment lexicon generation for an under-resourced language. Int. J. Comput. Linguistics Appl., 5(1):59-72, 2014. 6

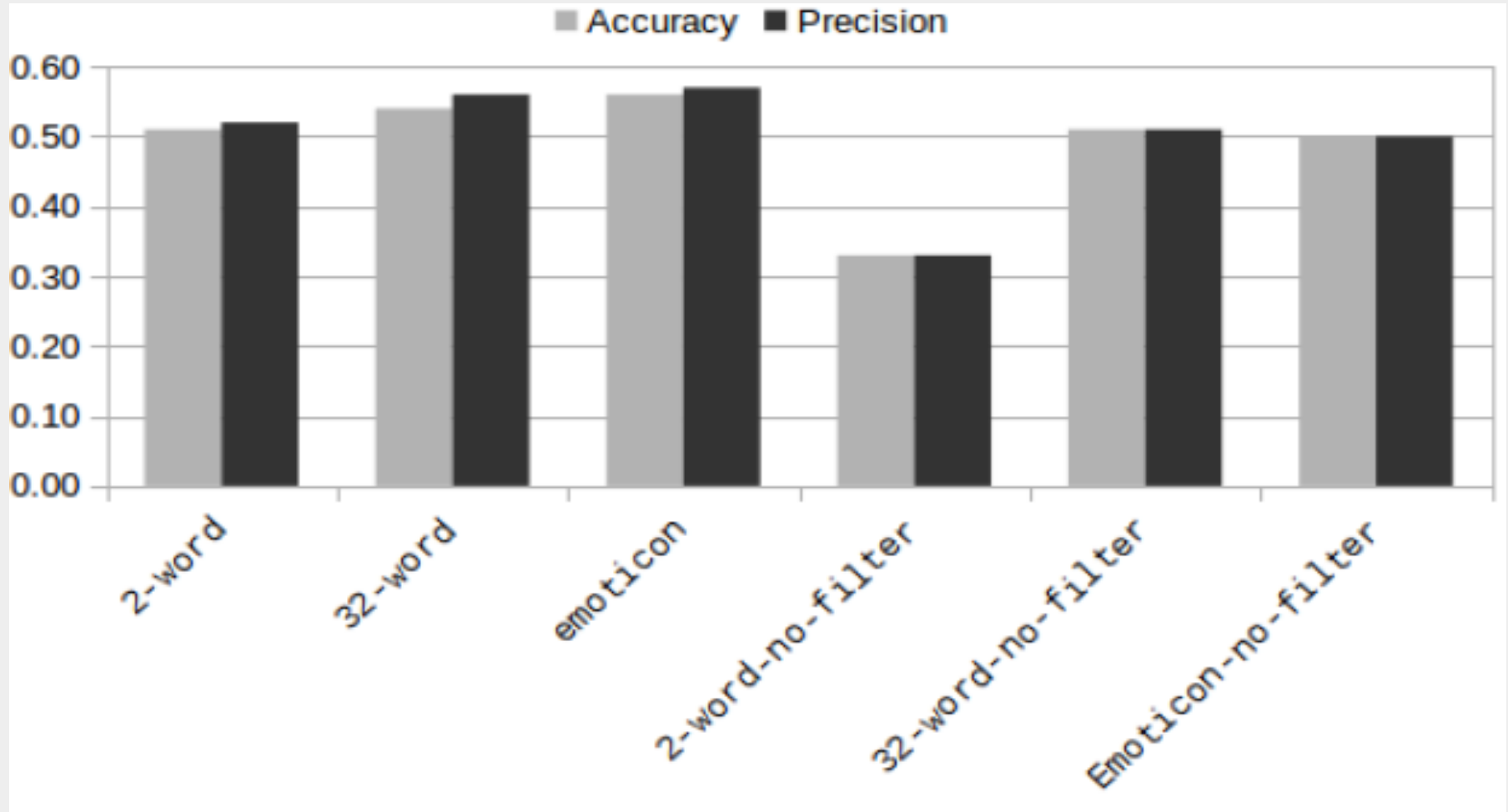


Questions



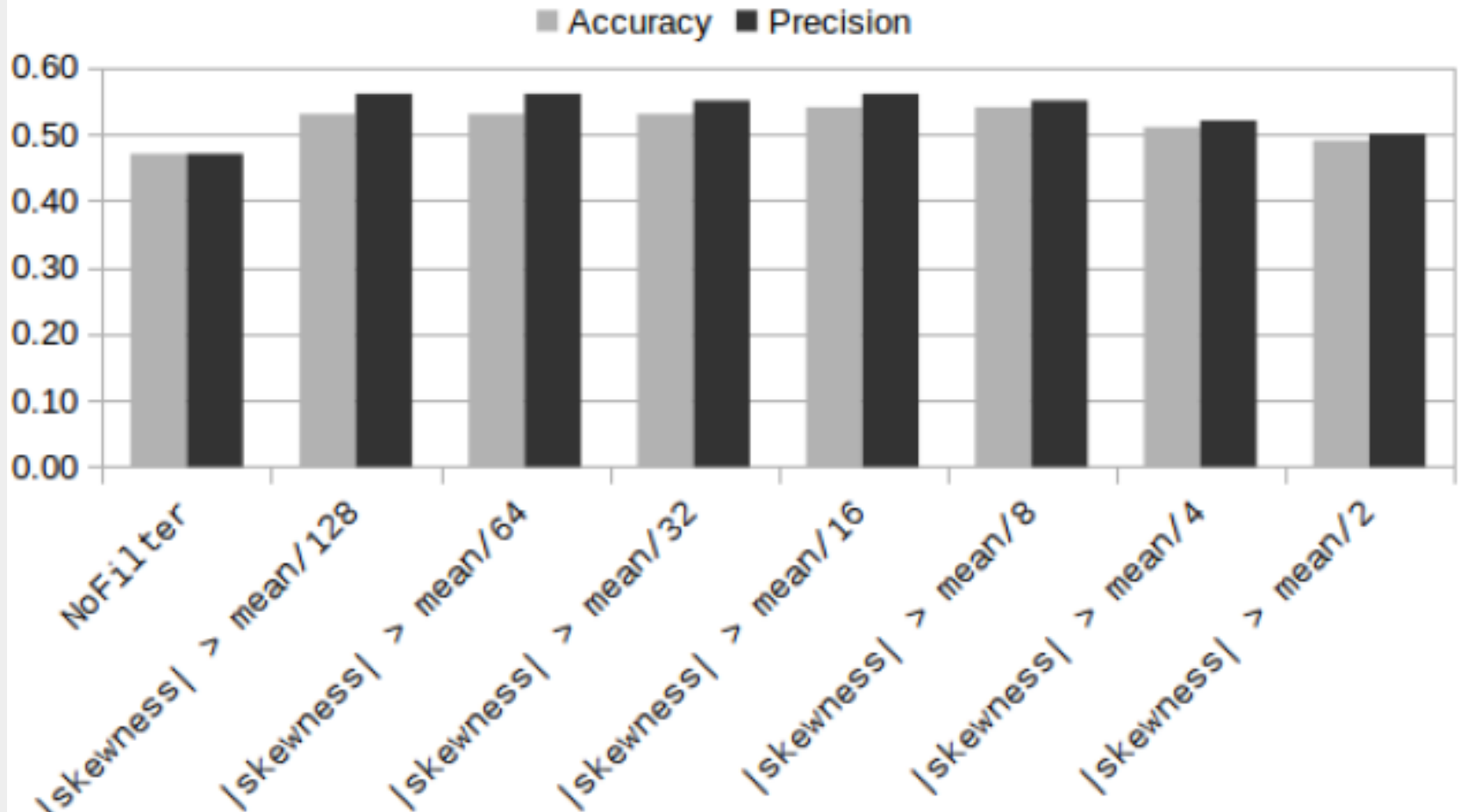
Build Filters

❖ EFFECT OF FILTERS



Build Filters

❖ SKEWNESS

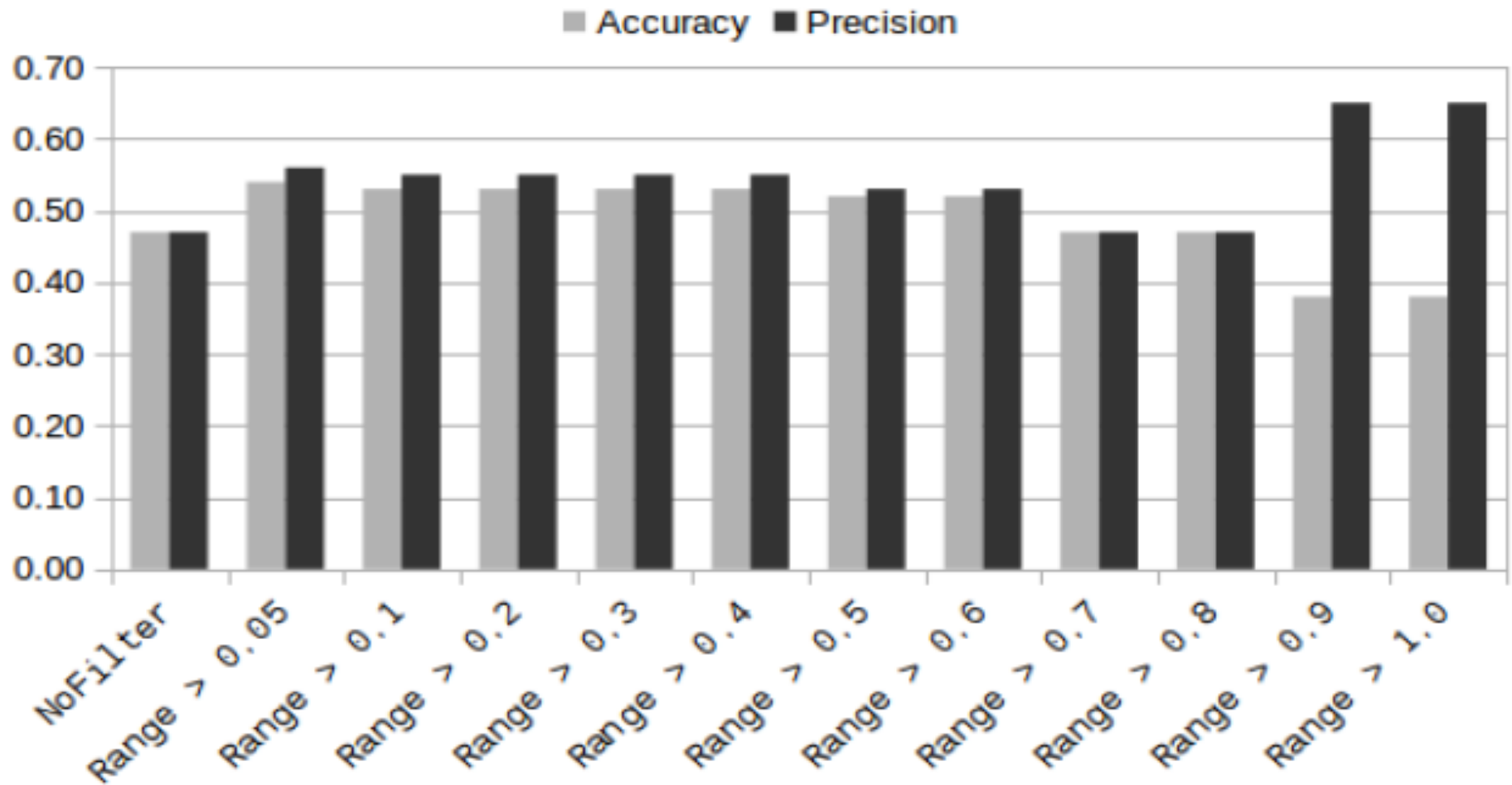




Build Filters

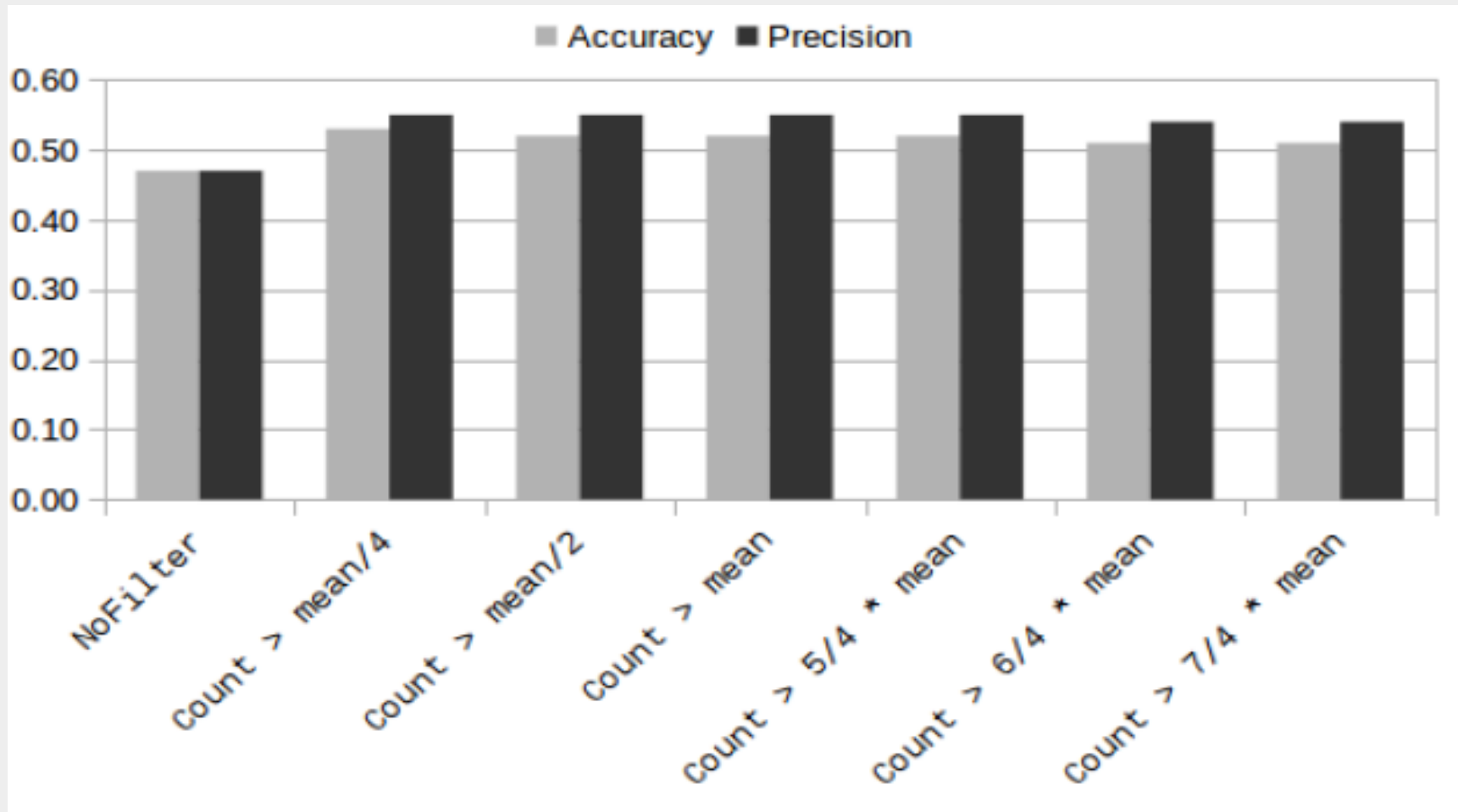


RANGE



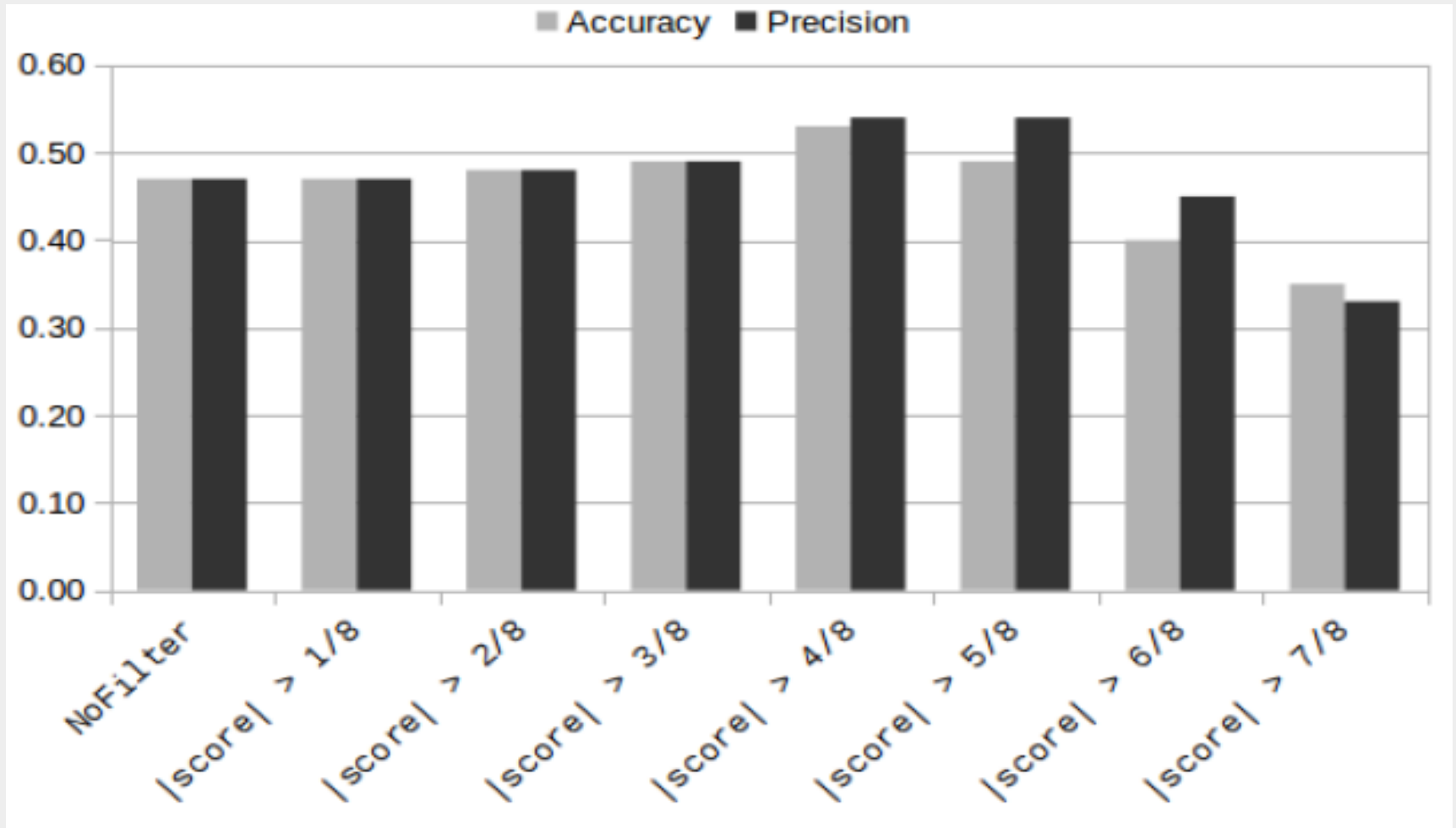
Build Filters

❖ COUNT



Build Filters

❖ SCORE





Unigrams (2-word Seed)

काठमाडौंमा बसितारबादी
हजारको लौरो उतारी अन्तरमिदेउवाले
चनियि शताब्दीमा जन्मएका पोता
बखिण्डन खोलनुहोस रगतले सक्छौं
समतिको महाकालि संकटमा
भुक्न मोदीको जातयिताको
गर्दैछन् मोर्चाको टुक्राउन
आजपर्न भत्काएर स्वकार
टपिर भत्काएर भारत सामन्त
जुवानेन्द्रको रतान जनसंख्याको
हुतहारा डरलाग्दो आवश्यक
इत्यादि वपिरति पश्चिमा गद्दारी
करीब पहाडमा गद्दारी
एभुलाई साहसियलाई कालापानी तैयार फास्ट
आक्यो गठबन्धनको राजसंस्था
गर्छस मधेसीलाई सम्पूर्णको संवधानको
भन्नि जातका कुकर्म
चरिका भनि परक्रियामा सटिमोल
मुटेको बपिपा युध परक्रियामा सटिमोल
गोरुको नबन्ने अधिकारका
दविगत जग्मा पीडितका
लनुपर्छ डराउने मोदीले संवधानक अन्तर्लगा
कालापानि प्रविधान गरनुपर्छ तमिले
आदधिस त्रि दान्त्य सासम संवधानमा
ग्रलाई कोशी लेनडुप बराबरी चीर
कोसी मन्त्रालय हस्ताक्षर उखाने
पराएको पजेरो नरिपेक्षता कार्यकरताहु
ढाल्ने अतकिर्मण समयसम अमुल मधेसमा जोषा
अमीहको छुन समयसम अमुल मधेसमा जोषा
संबधानको कटौती भाउमा सम्बन्धान

दाईको लक रेखाजी सुन्दरता हजुरहुको
छक्का लभ भायुक् परस्तुति मैसाप
नायिका रुद्रपरया हिममतवाली अर्त फलिमको हर्
खुसीको वजिया सुगन्ध हजुरहु सुनमा
फूल रसिभकामना गोबनिदे अग्रमि
लाईम युवती फलिम एल उत्तरोत्तर
बधाइ दामि सलाम परमाणति सरापेको बी
नपायको कवति अडयो मायाले हामिमत
सुन्दर नमस्ते बुक वधाई पन्जा
सेक्सी सुन्तली सफलता रमाइलो बहानी
बाइ प्रयिका हजुरको गुलमी सुन्दैछु मेरी
शुभ सुभ ठमिलाई काली रामप्यारी नमस्कार
मसि फलिमी हमाल हजुर परस्तुती
ओ हेर्छु प्यारी गधिरमुस वेस्ट सुपर
सवालको उहालाई राम्रो साझासवाल
बा दपि जतिको रोल हटि मालाले झुर
मंगलमय गायक हटि मालाले गुड
धन्यबाद मीठो हजुरको फ्यान दशमी
परस्तोता रमैलो आभार हाहाहा
परस्तोता रमैलो हर्दिय
सवालले मुरी हजुरलाई नारी घमिरी लगछ
बधिया दामी खुमे स्टार सफलताको
लाइक गा कहानी मार्मिक सुपार सुवाल
खचैर सेवाले बजिया ल परगतको
रेखाजी अमनिय रेखाको सतलि दीदी हरी
उपेक्षायमा हेरथिो सर धन्याबाद धननेबाद
सुर बदाई सवाल नगरेकै फोटो सोचाई
फलमि कार्यकालको सेल्फी रामरी रुद्र
हैत सुन्दै फलिममा रबनिदर अपलोड
श्रोता हरविंशबैनी दशमीको ददि आहा
रेखाले



Unigrams (32-word Seed)

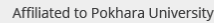
डकुटी आफैँ धुइका बखिण्डनकारी जानुन बकिने प्रचंड पल्केका लेन्डुप खाडल कलिको भट चनिया सुस्ता छस भाउमा नरिपेक्षता टिकापुर मालाले रतियाउने कालापानी प्रचन्डे पडकाउने तँलाई पडकाउने तँलाई कमशिन अंगकित रतितो भारुको ज्ञानेन्द्रको लुटेको यातना चुरोट सामन्त भुस्याहा धोतकी मोदकी डराउने लौरो मोदकी डराउने मचचाएर घाँटी गरछस गन्डकी चोरेर तमिहुरो असेल पसेर बेचदा जरे थुक्का कौडकी केरा थुक्का सुसलि महाकाल टिकपुर घुस्योहा पाखन्डी प्रभुलाई खोजेको पीडतिका नकिले खुमे भारत अस्थतिव खुमे

पुस्तनिधिहिरु रुद्रपरया मुदमा सुनिध घरलाई रमेलो साहित्यिक मरमाहत नषिपक्ष वजिया सटार रेख शधिर मुदलाई बहनि समेटन टमिलाई गोकर्ण उततरोत्तर समपुरणको लाईम प्रस्तोता जलिलाबाट गर्दछ सुनतली पूरणा लीको नमसते गराउनको भावपूर्ण सुपार साधर प्यारी बनिमूर मायाले कविता दाइको सन्तपत हरदिके अनुष्ठान मकवानपुर गुरुभएको शब्दमा पाईयो प्रस्तुति मलियोस आत्माको जविनमा सुन्दरी सुनमा बघाइ आबद्ध माता सुभ लाभको हमाल गुरु पृथ्वी मशिर हार्दीक समवेदना सुन्दरता चीर प्रशनको अमनिय श्रीकृष्ण शुभ राम्री कार्यक्रमले मातामना अक वैश्य जन्मदैनिको रामप्यारी समेटने रखाउने नमन जतिको उतारी साहित्यलाई सवालको गजल गरदछ सुन्दर दाइको फोटो बजिया सरको फलिमी भावपूर्ण सुभकामना आहा बधैछ मनपर्ने र शि समझनु दानी साझासवाल वयबल बधैछ दपि परिवारजनमा मैसाप सन्ध खोलनुहोस बखिण्डन ठजुरलाई अझसम्म गा साहित्य संगीत हेरयो वडा दशमीको लक रवदर नतिान्त बषिट धन्यवाद वविकशलि गहन वधाई स्वस्थ हामिमत श्रेष्ठ दशमी सुन्दैछ छुने पशुपतिनाथ लाईक भनि अम्बर चरि बनाँ रुद्र सफलता श्रोता भनि अम्बर चरि कार्याकालको वेस्ट छनोट आचार्य सुखद उहाको नारायण प्रगतिको भएकामा रहोस् प्रेरणा सुखद उहाको सेवाले मंगलमय समवेदना कुराँट छनोट पराएको हजुरलाई



Unigrams (Emoticon Seed)

दलालको बाटो गल्ली केई आर कालोबजारी पउने बाने गल्लि तरिने पसेर लगायो मुज्जा बर्ना मल्लो सफलता शधिर अम्बर शेखा अमनिय
जाडोमा बाउले खुमे खान्छन मकि पयारी हमिमत गि नडिर हातहु गोकर्ण चनिहमा
आउथयो औषध छोडौं ग्यासको फुरती गरसि तेरो सुगन्ध समझनु दामी प्रीय व्यक्त रहोस् भएकीमा जीको
सुकुम्बासी बलात्कारी बधि सुनको रामप्यारी सुर्ख सधिर सुखदे जयुले दाड लालसलाम
बबिरण पचेको हाहाकार आजबिन दोरा जन्मदनिको प्रगतिको जयुले सन्तपूत
भारुको भत्काएर धोतकी साग आत्माको अभविदन महाइयास् चाल्नु थापा
गन्डकी गुजारा पीडितिका थुक्क टाउकोमा खलती लोड पजेरो मारदै हेरछु प्रसारति गजल मुटुलाई
समार्ट सले लेनको छोक साहित्य बन्ने कार्यक्रमको साहलियकि घले धन्यवाद वेस्ट आबद्ध
चरिको दरबारमा थुक भर्न भाउमा श्रृंक्षण हजुरहुको प्रस्तनिधिहिरु परवारमा मालाले सूपार
येत्रो आर पसाब कागलाई खा थोक उतारी हजुरहुको प्रस्तनिधिहिरु परवारमा मालाले सूपार
लाउडा सुत्केरी पाइप लेनडप मरी हेलकिप्टर खोलनुहोस आभार सदा लाईक गर्दछौ
कारबाहा चरिको बकिने बेल बजिग कालापाना दशमीको एवम् चरि जतिको री नव वजिय
मारी लौरो रतितो लोकल मारी श्रद्धाान्जल वजिया दविगत
बदिशा भनदा काकरो सरकार धक्कार बेचि श्रद्धाान्जल वजिया दविगत
पाच उचालने बनाउछने धोस्त तेरो हाँडो तमिहुले भाडा बुझनुहोस दार्दिक सनदेस गरदछु दामा
कौडिको खादा खोजेको बाउको गरछस पावन हरदम श्रद्धा गरदछु मरमाहत टीम
लुकाउन गरछन पकाउने पाखन्डी डराइमर योगेश श्रद्धाान्जल अठोट सवाल
थुङ्कक ओभानो खाते पेटरोल कलकारखाना अतुतोपतुते समुन लालबबु सीता रमैलो हरदकि सुभ यात्रामा
बुडी जानेको बलत्कारी सरम जानेका चुसन् गरदछु चीर टमिलाई बहानी बनिमर
पचिछे लुटने पाता रतियाउने थुक्का लक सुनमा माता आजपन मुटुदालाई
फास्ट चढे बरिता चुसने लुटेको तलाई धोर्त सवालको श्रेष्ठको शान्तीको घडीमा रुदरप्रिया
हान्नु तराइमा पुजारी हडताल हान्नु तमिहुलाइ गोर्को सुखलाई आचार्य समृद्धिको सुर्य
भैहाल्यो घरलाई किन मान्छन मुटान तमिहुलाई कानै काटोको बगिनस्थल काँसा मोटर पडितिका मतले प्रस्तोता काँग्रेसका गनुहोला कार्यक्रमहरु अनुरम वविकशील

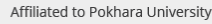


Page 46



Bigrams (32-word Seed)

मा.बहुमत का.दलाल चारो देशको दुकुटी आखामा.छारो देशको दुकुटी सफल.होस सुभकामना.छ माया.गरनु
गु.खाने भारत.लाई मा.भारतीय राजदुत.लाई सफा.गरने सुभकामना.छ माया.गरनु
गर्ने.काङ्ग्रेस तातो.न पनी.शुभकामना छोटी.समय बजिया.दशमी मदन.दाइ जनम.दनिको
सकदै.नौ.भने पनी.भारत सुपार.स्टार नायकि.रेखा थालनी.गर्नु परकोट.गर्दछु
महेन्द्र.ले भारतलाई.बेचेर नरिवाचति.कार्यकारी आँसु.दुख जय.श्री
दुख.दनि मालाले.स्वागत माथिल्लो.कर्णाली पेज.लाईक गरु.भएकोमा
मान्छे.मारेर यो.फिल्म कार्यकालको.शुभकामना राम्रो.मान्छे लाभ को सुख.शान्ति
पानी.माथिको एक.प्रदेश राखे.पनी कुकुर.हो मुद्रा.मा.बस्न
काङ्ग्रेस.एमाले रगत.को साले.चोर र.तनिका नाटक.हो र.पसनि धर्म.नरिपेक्षता
भारतलाई.गाल साला.चोर नेता.भनौदा काङ्ग्रेस.लाई मदन.भण्डारीको मधेसी.लाई जस्ता.पाता
अनि.नेता.क भारत.बरीधा कुकुर.ले दलाल.र पनी.भने मा.करोड पनी.भने
नाम.बेचेर लुट्ने.लुटेरा र.मधेशी भन्ने.बेर भर्षटाचार.गरने तेल.हालेर भर्षटाचार.गर्
आन्दोलन.मा पीडितका.लागा लाज.लागनु लाज.लागनु सला.कुकुर
घुस.खाने मधेश.मादलाल.इन कुकुर.को चोर हुनु सान्त्वना.गर्ने साले.कुकुर
एकातरि.कुम्लो नयिम.पालना धोती.को कुकुर.को चोर हुनु सान्त्वना.गर्ने साले.कुकुर
को.गुलामी र.भारतको यो.कुकुर बदिशी.को जनताको.आखामा को.दलाल धर्मिकार.छ
मन्त्री.पद सबै.थोक नारायण.शरेशूठ धेरै.बैछ धेरै.मन मन.पराएको को.सुभकामना
कौडको.भाउमा जेल.हालेर को.भन्छ र.चोट धेरै.शुभकामना मेवारी.कर्मिटर सरास.कार्यकारीको चारो.हुने



यो. नाटक ^{हनुमन् लख} बदिशी. को गरन. थाले तेरो. बाउ ^{मुनि. कवि} मन. पराएको वयक्त. गर्दछु ^{पुनः फर्क}
दलाल. को ^{जस्ता पेडा} को. दलाल ^{खाने. र} भो. भनेर ^{कुनै. कवि}
महेन्द्र. ले ^{अनुराग र} को. दलाल ^{खाने. र} भो. भनेर ^{कुनै. कवि}
जनता. माथ ^{हमिमत. छ} नसक्ने. ले ^{छोभा. गर्न} को. दलाल ^{खाने. र} भो. भनेर ^{कुनै. कवि}
गो. कुकुर ^{कसरी. कसरी} आफ्नो. पेट ^{सुख. चिन्ता. छैन} सु. भयो ^{राजदुत. लाई}
एक. छाक ^{मुकम्प. पीडातिका} सबै. नेताहरु ^{सबै. नेताहरु}
हाम्रा. नेताहरु ^{हाम्रा. नेताहरु} कौडको. भाउमा ^{साले. चोर}
खाने. हो ^{भुक्तम. पीडातिका} भारतलाई. बेचेर ^{सबै. नेताहरु}
र. पसनि ^{जस्ता. पाता} सोझा. जनता ^{माथेर. छानबनि. गर्ने}
लाई. नेपालको ^{आखामा. छारो} भनेको. यह ^{फाक. कल्लु}
देश. बेचेर ^{कौ. टाउको} का. दलाल ^{घुस. खाने} घुस. खाने ^{कौ. टाउको}
कानै. चरिको ^{तरिको. कर} तरिको. कर ^{के. गरछो} के. गरछो ^{पसो. पसु}
अतत्तोपत्तो. छैन ^{अतत्तोपत्तो. छैन}
चोर. लाई ^{चोर. लाई} चोर. डाका ^{बदिश. भ्रमण} बदिश. भ्रमण ^{जोगी. आए}
जोगी. आए ^{मानछे. मारेर} मानछे. मारेर ^{मानछे. मारेर} सबै. थोक ^{मानछे. मारेर}
गरी. लाख ^{गरी. लाख} के. नाला ^{गरी. लाख} के. नाला ^{गरी. लाख}
को. गुलामी ^{को. गुलामी} को. गुलामी ^{को. गुलामी}
देश. बेचने ^{देश. बेचने} मा. बहुमत ^{मा. बहुमत} पान्छ ^{पान्छ}
देशमा. रोजगार ^{देशमा. रोजगार} देखि. लाख ^{देखि. लाख}
सबै. दल ^{सबै. दल} भारतलाई. गर्ल ^{भारत. मारा}
रगत. को ^{रगत. को} पनी. संबिधान ^{पनी. संबिधान}
देश. बिगरने ^{देश. बिगरने} लात. हानेर ^{लात. हानेर}
कुकुर. ले ^{कुकुर. ले} कुकुर. ले ^{कुकुर. ले}
सुरक्षा. गर्न ^{सुरक्षा. गर्न} कत. हो ^{कत. हो}
चाल. हो ^{चाल. हो} यो. कुन ^{यो. कुन} लाज. छैन ^{लाज. छैन}
आखामा. छारो ^{आखामा. छारो}