**000**

**001**

**002**

**003**

**004**

**005**

**006**

**007**

**008**

**009**

**010**

**011**

**012**

**013**

**014**

**015**

**016**

**017**

**018**

**019**

**020**

**021**

**022**

**023**

**024**

**025**

**026**

**027**

**028**

**029**

**030**

**031**

**032**

**033**

**034**

**035**

**036**

**037**

**038**

**039**

**040**

**041**

**042**

**043**

**044**

**045**

**046**

**047**

**048**

**049**

**NAACL-HLT 2019 Submission \*\*\*. Confidential Review Copy. DO NOT DISTRIBUTE.**

**BHAAV (भाव) - A Text Corpus for Emotion Analysis from Hindi Stories**

**Anonymous NAACL submission**

**Abstract**

In this paper, we introduce a new Hindi text corpus (BHAAV - which means ‘emo- tions’ in Hindi) for analyzing emotions that a writer expresses through his characters in a story, as perceived by a narrator/reader. The corpus consists of 20,304 sentences col- lected from 230 different short stories span- ning 18 genres. Each sentence has been annotated by three native Hindi speakers with at least ten years of formal education in Hindi, with the goal of identifying one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. To our knowl- edge, this is the first and largest annotated text corpus for emotion analysis of stories in a low-resource language like Hindi. This paper discusses the scope of the corpus and its possible uses. We also provide a detailed analysis of the dataset and train baseline classifiers reporting their perfor- mances. The dataset will be made publicly available.

**1 Introduction**

Emotion analysis from text is the study of identifying, classifying and analyzing emo- tions (e.g., *joy*, *sadness*) as expressed and reflected in a piece of given text (Yadollahi et al., 2017). It’s wide range of applications in areas such as - *customer relation manage- ment* (Bougie et al., 2003), *dialogue systems* (Ravaja et al., 2006), *intelligent tutoring sys- tems* (Litman and Forbes-Riley, 2004), *analyz- ing human communications* (Kövecses, 2003), *natural text-to-speech systems* (Francisco and Gervás, 2006), *assistive robots* (Breazeal and Brooks, 2005), *product analysis* (Knautz et al., 2010), and *studying psychology from social me- dia* (De Choudhury et al., 2013), has drawn considerable attention from the scientific com-

munity making it one of the important areas of research in computational linguistics.

munity making it one of the important areas of research in computational linguistics.

munity making it one of the important areas of research in computational linguistics.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

Almost all the methods and resources devel- oped in this domain deals with English lan- guage (Yadollahi et al., 2017), making our understanding of expression of emotions only limited to English text. This paper describes our attempt to develop a text corpus for emo- tion analysis from stories written in Hindi, which is one of the 22 official languages of In- dia and is among the top five most widely spo- ken languages in the world1. Despite its wide usage, there are no text based resources for emotion analysis in Hindi, making the resource shared in this work as the first and largest annotated corpus for studying emotions from Hindi text, and facilitating development of lin- guistic resources in low-resource languages.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

According to a joint report by KPMG and Google2 published in 2017, there are 234 mil- lion Internet users in India using one of the Indian languages as their medium of communi- cation against 175 million users using English. This gap is predicted to increase by 2021, with users using Indian languages reaching 536 mil- lion. Among the 122 languages officially recog- nized by Indian constitution, Hindi is the most spoken, followed by Bengali and Telugu. This has also led to a sudden rise in interest from so- cial media companies like Facebook, and Inter- net search companies like Google to increase their support for popularly used Indian lan- guages, making our work in this domain apt and timely.

Related to the task of emotion analysis

Related to the task of emotion analysis

1https://en.wikipedia.org/wiki/List\_of\_ languages\_by\_number\_of\_native\_speakers

1https://en.wikipedia.org/wiki/List\_of\_ languages\_by\_number\_of\_native\_speakers

1https://en.wikipedia.org/wiki/List\_of\_ languages\_by\_number\_of\_native\_speakers

2https://assets.kpmg.com/ content/dam/kpmg/in/pdf/2017/04/ Indian-languages-Defining-Indias-Internet.pdf

2https://assets.kpmg.com/ content/dam/kpmg/in/pdf/2017/04/ Indian-languages-Defining-Indias-Internet.pdf

2https://assets.kpmg.com/ content/dam/kpmg/in/pdf/2017/04/ Indian-languages-Defining-Indias-Internet.pdf

2https://assets.kpmg.com/ content/dam/kpmg/in/pdf/2017/04/ Indian-languages-Defining-Indias-Internet.pdf

**050**

**050**

**051**

**051**

**052**

**052**

**053**

**054**

**055**

**055**

**056**

**056**

**057**

**058**

**059**

**060**

**061**

**061**

**061**

**062**

**062**

**062**

**063**

**063**

**063**

**064**

**064**

**064**

**065**

**065**

**065**

**066**

**066**

**066**

**067**

**067**

**067**

**068**

**068**

**068**

**069**

**069**

**069**

**070**

**070**

**070**

**071**

**071**

**071**

**072**

**072**

**072**

**073**

**073**

**073**

**074**

**074**

**074**

**075**

**075**

**075**

**076**

**076**

**076**

**077**

**077**

**077**

**078**

**078**

**078**

**079**

**079**

**079**

**080**

**080**

**080**

**081**

**081**

**081**

**082**

**082**

**083**

**083**

**083**

**084**

**084**

**084**

**085**

**085**

**085**

**086**

**086**

**086**

**087**

**087**

**087**

**088**

**088**

**088**

**089**

**089**

**089**

**090**

**090**

**090**

**091**

**091**

**091**

**092**

**092**

**092**

**093**

**093**

**093**

**094**

**094**

**094**

**095**

**095**

**096**

**096**

**096**

**097**

**097**

**097**

**097**

**098**

**098**

**098**

**099**

**099**

**099**

1

**100**

**101**

**102**

**103**

**104**

**105**

**106**

**107**

**108**

**109**

**110**

**111**

**112**

**113**

**114**

**115**

**116**

**117**

**118**

**119**

**120**

**121**

**122**

**123**

**124**

**125**

**126**

**127**

**128**

**129**

**130**

**131**

**132**

**133**

**134**

**135**

**136**

**137**

**138**

**139**

**140**

**141**

**142**

**143**

**144**

**145**

**146**

**147**

**148**

**149**

**NAACL-HLT 2019 Submission \*\*\*. Confidential Review Copy. DO NOT DISTRIBUTE.**

in Hindi, previous attempts have been made in developing corpus for predicting emotions from Hindi-English code switched language used in social media (Vijay et al., 2018) (2,866 sentences) and from auditory speech signals (Koolagudi et al., 2011). Some work has been undertaken in a closely related task of senti- ment analysis and datasets have been created for identifying sentiments expressed in movie reviews (Mittal et al., 2013) (664 reviews), Hindi blogs (Arora, 2013), and generating lex- ical resources like Hindi Senti-Wordnet (Joshi et al., 2010). Given the dearth of resources for analyzing emotions from Hindi text, we take this opportunity to present and publicly share ‘BHAAV’ - a corpus of 20,304 sentences col- lected from 230 different short stories span- ning across 18 genres, written in Hindi. Each sentence has been annotated by three native Hindi speakers with at least ten years of formal education in Hindi, with the goal of identify- ing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*.

Stories are melting pot of different types of emotions expressed by the author through the characters and plots that he develops in his writing. Emotions in storytelling has been pre- viously studied resulting in identification of six basic types of emotional arcs in English stories (Reagan et al., 2016). This motivated us to de- velop BHAAV from Hindi stories. We believe that apart from studying emotions in Hindi text, the presented corpus would also enable studies related to the analysis of Hindi litera- ture from the perspective of identifying the in- herent emotional arcs. It also has the potential to catalyze research related to human text-to- speech systems geared towards improving au- tomated storytelling experiences. We keep the order of the sentences intact as they occur in their source story. This makes the corpus ideal for performing temporal analysis of emotions in the stories, and provides enough informa- tion for training machine learning models that takes into account temporal context.

Major contributions of our work are: **-** *Publicly share the first and the largest an- notated corpus of 20,304 sentences in Hindi (BHAAV), collected from 230 different short stories spanning across 18 genres, identifying one of the five different emotions - anger, joy,*

*suspense, sad, and neutral.* **-** *Describe potential applications of BHAAV corpus, the process of annotation and main challenges in creating an emotion analysis text corpus in a low-resource language like Hindi.* **-** *Report performances of baseline classifiers trained for identifying emotion expressed in a sentence of a story written in Hindi.*

*suspense, sad, and neutral.* **-** *Describe potential applications of BHAAV corpus, the process of annotation and main challenges in creating an emotion analysis text corpus in a low-resource language like Hindi.* **-** *Report performances of baseline classifiers trained for identifying emotion expressed in a sentence of a story written in Hindi.*

*suspense, sad, and neutral.* **-** *Describe potential applications of BHAAV corpus, the process of annotation and main challenges in creating an emotion analysis text corpus in a low-resource language like Hindi.* **-** *Report performances of baseline classifiers trained for identifying emotion expressed in a sentence of a story written in Hindi.*

*suspense, sad, and neutral.* **-** *Describe potential applications of BHAAV corpus, the process of annotation and main challenges in creating an emotion analysis text corpus in a low-resource language like Hindi.* **-** *Report performances of baseline classifiers trained for identifying emotion expressed in a sentence of a story written in Hindi.*

*suspense, sad, and neutral.* **-** *Describe potential applications of BHAAV corpus, the process of annotation and main challenges in creating an emotion analysis text corpus in a low-resource language like Hindi.* **-** *Report performances of baseline classifiers trained for identifying emotion expressed in a sentence of a story written in Hindi.*

*suspense, sad, and neutral.* **-** *Describe potential applications of BHAAV corpus, the process of annotation and main challenges in creating an emotion analysis text corpus in a low-resource language like Hindi.* **-** *Report performances of baseline classifiers trained for identifying emotion expressed in a sentence of a story written in Hindi.*

*suspense, sad, and neutral.* **-** *Describe potential applications of BHAAV corpus, the process of annotation and main challenges in creating an emotion analysis text corpus in a low-resource language like Hindi.* **-** *Report performances of baseline classifiers trained for identifying emotion expressed in a sentence of a story written in Hindi.*

*suspense, sad, and neutral.* **-** *Describe potential applications of BHAAV corpus, the process of annotation and main challenges in creating an emotion analysis text corpus in a low-resource language like Hindi.* **-** *Report performances of baseline classifiers trained for identifying emotion expressed in a sentence of a story written in Hindi.*

*suspense, sad, and neutral.* **-** *Describe potential applications of BHAAV corpus, the process of annotation and main challenges in creating an emotion analysis text corpus in a low-resource language like Hindi.* **-** *Report performances of baseline classifiers trained for identifying emotion expressed in a sentence of a story written in Hindi.*

**2 Related Work**

**2 Related Work**

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

In this section, we present works relevant to that of ours. It is necessary to mention that there has been extensive work in sentiment analysis especially in the past two decades. Although, there is significant intersection be- tween techniques used for sentiment analysis and emotion analysis, yet the two are different in many ways. Emotion analysis is often tack- led at a fine grained level and has historically proved to be more challenging due to subtleties involved in identifying and defining emotions that are orthogonal to each other. Addition- ally, resources for emotion analysis are scarce when compared to sentiment analysis. For a detailed survey of methods, datasets and the- oretical foundations on sentiment analysis and emotion analysis, please refer (Yadollahi et al., 2017; Lei et al., 2018; Cambria et al., 2017; Po- ria et al., 2017)

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

Analyzing emotions from text has been pri- marily manifested through four different types of tasks - *Emotion Detection* (Gupta et al., 2013), *Emotion Polarity Classification* (Alm et al., 2005), *Emotion Classification* (Yang et al., 2007), and *Emotion Cause Detection* (Gao et al., 2015). The scope of this work is limited to the task of *Emotion Classification*. (Pang et al., 2008) mentions that emotions are expressed at four levels - *morphological*, *lexi- cal*, *syntactic* and *figurative*, and noted that as we move from morphological to figurative, the difficulty of the emotion analysis task in- creases and number of resources for the same decrease. Developed from stories written in a morphologically rich language, BHAAV pri- marily deals with the first and the last levels. Most of the work for creating data resources for emotion analysis has been fairly limited to building emotion lexicons (Strapparava et al., 2004; Pennebaker et al., 2001; Shahraki and Zaiane, 2017; Mohammad and Turney, 2013),

2

**150**

**150**

**150**

**151**

**151**

**151**

**152**

**152**

**152**

**153**

**153**

**153**

**154**

**154**

**154**

**155**

**155**

**155**

**156**

**156**

**156**

**157**

**157**

**157**

**158**

**158**

**159**

**159**

**159**

**160**

**160**

**160**

**161**

**161**

**161**

**162**

**162**

**162**

**163**

**163**

**163**

**164**

**164**

**164**

**165**

**165**

**165**

**166**

**166**

**166**

**167**

**167**

**167**

**168**

**168**

**168**

**169**

**169**

**169**

**170**

**170**

**170**

**171**

**171**

**171**

**172**

**172**

**172**

**173**

**173**

**173**

**174**

**174**

**174**

**175**

**175**

**175**

**176**

**176**

**176**

**177**

**177**

**177**

**178**

**178**

**178**

**178**

**179**

**179**

**179**

**180**

**180**

**180**

**181**

**181**

**181**

**182**

**182**

**182**

**183**

**183**

**183**

**184**

**184**

**184**

**185**

**185**

**185**

**186**

**186**

**186**

**187**

**187**

**187**

**188**

**188**

**188**

**189**

**189**

**189**

**190**

**190**

**190**

**191**

**191**

**191**

**192**

**192**

**192**

**193**

**193**

**193**

**194**

**194**

**194**

**195**

**195**

**195**

**196**

**196**

**196**

**197**

**197**

**197**

**198**

**198**

**198**

**199**

**199**

**199**

**200**

**201**

**202**

**203**

**204**

**205**

**206**

**207**

**208**

**209**

**210**

**211**

**212**

**213**

**214**

**215**

**216**

**217**

**218**

**219**

**220**

**221**

**222**

**223**

**224**

**225**

**226**

**227**

**228**

**229**

**230**

**231**

**232**

**233**

**234**

**235**

**236**

**237**

**238**

**239**

**240**

**241**

**242**

**243**

**244**

**245**

**246**

**247**

**248**

**249**

**NAACL-HLT 2019 Submission \*\*\*. Confidential Review Copy. DO NOT DISTRIBUTE.**

or concentrated in annotating emotions of in- dividual sentences without giving any con- text (Strapparava and Mihalcea, 2007). This approach, as indicated by many, is a non- holistic approach for a task such as emotion analysis (Schwarz-Friesel, 2015; Ortony et al., 1987). BHAAV not only presents annotated sentences, but also provides their context.

Lastly, when it comes to the task of analyz- ing emotions from text, there are no datasets available in Hindi. Although, resource-poor Indian languages have started catching up their richer counterparts in the domain of sen- timent analysis (Mittal et al., 2013; Arora, 2013; Joshi et al., 2010), yet sufficient work needs to be done considering the pace at which these languages are finding their uses in mod- ern digitally driven India. The lack of re- sources can be judged from the ‘wide’ usage of one of the very few Hindi datasets for sen- timent analysis task (Balamurali et al., 2012). It consists of just 200 positive and negative sentences for each of the two major Indian lan- guages, Hindi and Marathi. Another popular and a recent attempt is by (Patra et al., 2015). They released a dataset containing approxi- mately 1500 tweets for each of the languages of Hindi, Bengali and Tamil for Aspect Based Sentiment Analysis. BHAAV is certainly an attempt to fill this gap and create a large, ef- fective and high quality resource for emotion mining from text.

**3 Language Specific Challenges**

As already mentioned and pointed in (Yadol- lahi et al., 2017), the computational methods used in the tasks pertaining to sentiment anal- ysis can readily be applied to the emotion anal- ysis tasks. Therefore, the challenges for emo- tion analysis from text are very similar to that of the domain of sentiment analysis from text. For a detailed description of the challenges one can refer (Mohammad, 2017). However, our task of identifying emotions from sentences poses additional challenges due to the inherent characteristics of Hindi language. We point out some of these language specific challenges as identified by (Arora, 2013), in order to draw a complete picture of the intricacies of the task and emphasize on the fact that there is a scope of developing methods specific to Hindi, and

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

not all methods developed for English can be directly translated to Hindi. **Word Order** - The order in which words ap- pear in a sentence plays an important role for determining polarity as well as subjectiv- ity of the text. As opposed to English, which is a *fixed order language*, Hindi is a *free or- der language*. For any sentence in English to be grammatically correct the ‘subject’ (S) is followed by ‘verb’ (V), which is followed by ‘object’ (O) - [SVO]. For example the English sentence - *Ram ate three mangoes*, which fol- lows the [SVO] pattern, can be expressed in three ways in Hindi that do not adhere to the [SVO] pattern - ‘राम न तीन आम खाया’ [SVO], ‘तीन आम खाया राम न’ [OVS], and ‘खाय तीन आम राम न’ [VOS]. This lack of order can pose challenges to the machine learning algorithms that take into account the order of the words. **Morphological Variations** - Hindi language is morphologically rich. This means that a lot more information can be expressed in a word in Hindi for which one might end up writing many more words in English. One of the ex- ample is that of expressing genders. For exam- ple, when using the word ‘खायगी’, which means ‘will eat’ in English, one can not only indicate that someone will eat but also provide cues of the person’s gender (in this case female - the male variant is ‘खायगा’). **Handling Spelling Variations** - A word with the same meaning can appear with mul- tiple spelling variations. Occurrence of such variations can pose challenges for the machine learning models that has to take into account all the spelling variants. For example the word ‘मह गा’, which means ‘costly’ has another vari- ant महगा that means the same. **Lack of Resources** - The lack of lexicons, developed techniques and elaborate resources in Hindi also adds to the challenge, which is also one of the main motivations behind this work.

**4 Corpus Generation and**

**4 Corpus Generation and**

**4 Corpus Generation and**

**Annotation**

**Annotation**

**Annotation**

One of our primary aims was to create a man- ually annotated large corpus for performing emotion analysis from text in Hindi. We also wanted to capture the context in which a given piece of text occurs. Therefore, we decided to

One of our primary aims was to create a man- ually annotated large corpus for performing emotion analysis from text in Hindi. We also wanted to capture the context in which a given piece of text occurs. Therefore, we decided to

One of our primary aims was to create a man- ually annotated large corpus for performing emotion analysis from text in Hindi. We also wanted to capture the context in which a given piece of text occurs. Therefore, we decided to

One of our primary aims was to create a man- ually annotated large corpus for performing emotion analysis from text in Hindi. We also wanted to capture the context in which a given piece of text occurs. Therefore, we decided to

One of our primary aims was to create a man- ually annotated large corpus for performing emotion analysis from text in Hindi. We also wanted to capture the context in which a given piece of text occurs. Therefore, we decided to

One of our primary aims was to create a man- ually annotated large corpus for performing emotion analysis from text in Hindi. We also wanted to capture the context in which a given piece of text occurs. Therefore, we decided to

3

**250**

**250**

**250**

**251**

**251**

**251**

**252**

**252**

**252**

**253**

**253**

**253**

**254**

**254**

**254**

**255**

**255**

**255**

**256**

**256**

**256**

**257**

**257**

**257**

**258**

**258**

**258**

**259**

**259**

**259**

**260**

**260**

**260**

**261**

**261**

**261**

**262**

**262**

**262**

**263**

**263**

**263**

**264**

**264**

**264**

**265**

**265**

**265**

**266**

**266**

**266**

**267**

**267**

**267**

**268**

**268**

**268**

**269**

**269**

**269**

**270**

**270**

**270**

**271**

**271**

**271**

**272**

**272**

**272**

**273**

**273**

**273**

**274**

**274**

**274**

**275**

**275**

**275**

**276**

**276**

**276**

**277**

**277**

**277**

**278**

**278**

**278**

**279**

**279**

**279**

**280**

**280**

**280**

**281**

**281**

**282**

**282**

**282**

**283**

**283**

**283**

**284**

**284**

**284**

**285**

**285**

**285**

**286**

**286**

**286**

**287**

**287**

**287**

**288**

**288**

**288**

**289**

**289**

**289**

**290**

**290**

**290**

**291**

**291**

**291**

**292**

**292**

**292**

**293**

**293**

**293**

**293**

**294**

**294**

**294**

**295**

**295**

**295**

**296**

**296**

**296**

**297**

**297**

**297**

**298**

**298**

**298**

**299**

**299**

**299**

**300**

**301**

**302**

**303**

**304**

**305**

**306**

**307**

**308**

**309**

**310**

**311**

**312**

**313**

**314**

**315**

**316**

**317**

**318**

**319**

**320**

**321**

**322**

**323**

**324**

**325**

**326**

**327**

**328**

**329**

**330**

**331**

**332**

**333**

**334**

**335**

**336**

**337**

**338**

**339**

**340**

**341**

**342**

**343**

**344**

**345**

**346**

**347**

**348**

**349**

**NAACL-HLT 2019 Submission \*\*\*. Confidential Review Copy. DO NOT DISTRIBUTE.**

**Dataset/Genre Fleiss’s Kappa Krippendorff’s alpha** BHAAV dataset 0.80241357 0.802416004636 आदश वादी (Idealist) 0.741594061693 0.740462670362 मपरक (Romantic) 0.906633895329 0.90659267822 शहर जीवन (Urban Life) 0.825558745183 0.824972237422 शोषक और शो षत वग (Exploiter and Exploited Class) 0.798745458224 0.797991236844 नी तपरक (Moral Stories) 0.865927372214 0.865857301713 कसान जीवन (Life of a Farmer) 0.885919097027 0.886009070648 ऐ तहा सक (Historical) 0.921326884658 0.921310220449 रणादायक (Inspirational) 0.897278304968 0.897266011499 दश भ सभ धत (Patriotic) 0.991438728869 0.991446652425 गत जीवन क सम या (Personal Issues/Problems) 0.879354980254 0.879386098588 ढ़ और अध व ास (Dogmatic and Superstitious) 1.0 1.0 सय प रवार क सम या (Joint Family Problems) 0.887607633305 0.887631947099 रह यमयी (Mystery) 0.908185638589 0.908330457028 यथाथ वादी (Realistic and Pragmatic) 0.912201981538 0.912237344469 ामीण जीवन (Village Life) 0.967427620692 0.967434538508 उपदशपरक (Instructive) 0.919704200254 0.919700170635 भोग ए यथाथ क कहानी (Real Stories) 0.911361226137 0.911352174976 समाज सधारक (Society and its Reformation) 0.878218452096 0.878149647043

Table 1: Inter-annotator Agreements as measured us- ing Fleiss‘s Kappa (Fleiss and Cohen, 1973) and Krip- pendorff‘s alpha (Krippendorff, 2011) for the entire BHAAV dataset and for each genre.

extract all the sentences from short stories be- longing to genres popular in Hindi. We started with 30 genres, but narrowed down to only 18, depending on the availability of online content. Throughout the process of deciding on genres and finding online content relevant to them, we took help from an expert in Hindi litera- ture, who provided 500 online URLs contain- ing a popular short story belonging to one of the genres. Table 1, provides a list of gen- res available in our final dataset. Whenever possible we also searched for an audio book3 where the same story has been narrated by a narrator. This was done in order to help the annotators during the annotation process, in case they have to refer to examples of how a narrator/reader would express the emotion of a sentence in the context of the story. All our annotators were native Hindi speaking volun- teers who had a minimum of 10 years of formal education in Hindi, and showed great interest in reading the stories.

**Emotion Sample Sentences**

**joy** बादशाह न कहा त हार कहानी पहली दोन स अ धक मनोरजक ह (The king said that

your story is more entertaining than the previous two stories) **anger** पया नई दगा तो उसका खाल उतारकर बाजार म बच दगा (If he doesn’t gives the money then I will take

out his skin and sell it in the market) **suspense** मज र न अब तक तो झलक भर दखी थी अब तो उस पर नजर भर दखा तो ठगा सा खड़ा रह गया (Till now the

worker had only seen his glimpses, but when he saw him fully he was just stunned) **sad** उसन आस होत ए म मी क ओर दखा (With teary eyes he saw towards his mother) **neutral** म इसक मा (I am his mother)

Table 2: Sample sentences from BHAAV dataset for each emotion label.

All the URLs were scraped and the text was extracted from them. Not all of them could be retrieved. We ended up retrieving and extracting text from 230 stories. The ex-

3Example of audio books for some of the stories - https://www.youtube.com/user/sameergoswami/ playlists

tracted text was split into sentences in an au- tomated way and contained many unnecessary text that were not a part of the story. Dur- ing the annotation process the annotators fil- tered the unwanted text and only annotated the relevant portion. Whenever the sentences were not correctly split, the annotators also corrected them. A total of 5 annotators were used for annotating the entire corpus, such that each sentence gets at-least 3 annotations. During the annotation process the annotators had access to the main URL of the story and the list of audio books. Each story was anno- tated in one sitting and it took 9 months for finishing the process.

The guidelines for annotating emotions was designed to be very short and concise with re- gards to the definitions of the categories to be assigned. All the definitions were kept brief and aligned with (Plutchik, 1984), along with sample annotated sentences from the domain. We asked the annotators to identify only one of the five emotions expressed in a sentence of a story - *anger*, *joy*, *suspense*, *sad*, and *neutral*. We went with the above categories of emo- tions mainly due to their extensive use in other works and also included *suspense* as we were dealing with the domain of stories in which *suspense* is often a popular emotion infused by the authors in creating interesting plots. The annotators were instructed not to be biased by their own emotions towards a statement in the story while labeling them, and was asked to identify only the emotion that an unbiased narrator/reader of that story would like to ex- press while reading it to someone. Whenever confused, they were asked to refer the audio book of the story if available, or one of the authors, or mark it as *neutral* if that doesn’t clear the confusion.

**Emotion No. of Sentences No. of Sentences**

**(Train data)**

**No. of Sentences (Test data) joy** 2,463 2,242 221 **anger** 1,464 1,321 143 **suspense** 1,512 1,389 123 **sad** 3,168 2,843 325 **neutral** 11,697 10,478 1,219

Table 3: Distribution of sentences in different cate- gories of emotions in the BHAAV dataset.

General statistics of the dataset is presented in Table 3. The overall inter-annotator agree- ments and the agreements for individual gen- res are presented in Table 1. Some samples

4

**350**

**351**

**352**

**353**

**354**

**355**

**356**

**357**

**358**

**359**

**360**

**361**

**362**

**363**

**364**

**365**

**366**

**367**

**368**

**369**

**370**

**371**

**372**

**373**

**374**

**375**

**376**

**377**

**378**

**379**

**380**

**381**

**382**

**383**

**384**

**385**

**386**

**387**

**388**

**389**

**390**

**391**

**392**

**393**

**394**

**395**

**396**

**397**

**398**

**399**

**400**

**401**

**402**

**403**

**404**

**405**

**406**

**407**

**408**

**409**

**410**

**411**

**412**

**413**

**414**

**415**

**416**

**417**

**418**

**419**

**420**

**421**

**422**

**423**

**424**

**425**

**426**

**427**

**428**

**429**

**430**

**431**

**432**

**433**

**434**

**435**

**436**

**437**

**438**

**439**

**440**

**441**

**442**

**443**

**444**

**445**

**446**

**447**

**448**

**449**

**NAACL-HLT 2019 Submission \*\*\*. Confidential Review Copy. DO NOT DISTRIBUTE.**

sentences as annotated by the annotators are shown in Table 2. Next, we present some of the challenges that we faced during the annota- tion process that we think should be explicitly pointed out in order to provide a true picture of the corpus as well as to give an idea of the difficulties in carrying out such a process.

**4.1 Challenges in Annotation**

Apart from the challenge of annotating a low- resource language for which one can seldom get high quality crowd workers, there were cer- tain challenges that were both specific to the domain of stories as well as generic ones pe- culiar to the tasks of sentiment and emotion analysis. Some of the prominent ones as iden- tified from the feedbacks of the annotators are presented below with examples. **Identifying Implicit Emotions** - The an- notators were asked to identify the emotions whenever it was both explicitly and implicitly expressed. Identifying implicit emotions were sometimes confusing for the annotators and on taking a closer look we did find some of them being marked as neutral. An example of ex- plicitly expressed emotion would be - Exam- ple 1, in which the speaker by using the words such as सहावना (refreshing), मनोहर (beautiful) clearly indicates that he is happy with the na- ture, thus expressing his joy in the statements.

**.** *Example 1 -* कतना मनोहर, कतना सहावना भाव ह| व पर अजीब ह रयाली ह, खत म कछ अजीब रौनक ह, आसमान पर कछ अजीब ला लमा ह| *(It is such a beautiful and enjoyable feeling. There is a strange greenery on the trees, some strange liveliness in the fields, there is some weird but enjoyable redness in the sky)*

An example of implicitly expressed emotion would be - Example 2, in which a childns grandmother is complaining about her son be- ing too hasty of going to the mosque. She complains of his ignorance of knowing any- thing about driving a household and its in- herent difficulties. Although there aren’t any explicit word indicating her state of the mind, there is an implicit pointer that she is feel- ing irritated due to the haste and hence is an- gry over him. These types of emotions are totally contextual and could be identified only while reading the story. We believe that cap- turing these emotions are also necessary in order to make our annotation process holis- tic. Although, we don’t train any classification

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

model in this work that can take these types of context in order to predict the final emo- tion of a sentence, yet we think that BHAAV as a dataset provides an opportunity to build such contextual models making it a rich cor- pus unlike many other previous ones as already pointed out in Section 2. We would certainly like to take it up as a future work. **.** *Example 2 -* अब ज दी पड़ ह क लोग ईदगाह य नह चलत| इ ह गह थी क च ता स या योजन| *(Now he is feeling why don*n *t people go to the mosque a little faster. What do they (the children) know about household chores)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Primary Target of Opinion** - Another chal- lenge comes when there is not even an im- plicit clue in the immediate context of a sen- tence. For instance, in a story, sometimes a character is developed as an adversary to a particular prop (or, PTO (Primary Target of Opinion). The prop can be another charac- ter or some inanimate object or phenomena. From the start of the story, the character ex- presses his emotions in a characteristic man- ner towards that PTO. Thus if a sentence or a context does not have any explicit clues to know the state of the mind of the character, identifying the PTO and the character‘s emo- tions towards PTO gives some connotation to that sentence. This is in line to what was sug- gested in the work (Mohammad, 2016). An example of such an instance as presented in Example 3, can be derived from the famous story by Premchand, oEidgahp. The follow- ing sentence when read in isolation could po- tentially trick someone into thinking whether the boy speaking these dialogues is expressing mercy or even neutrality, when he is actually expressing joy. **.** *Example 3 -* मोह सन- ल कन दल म कह रह ह ग क मल तो खा ल | *(Mohsin- But in the hearts, they must be thinking that if they could get it, they would eat it)*

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

**Sarcasm** - A common challenge which anno- tators faced while annotating the dataset is the case of sarcasm, which is again prevalent in most of the previous works in sentiment and emotion analysis. Sarcasm, as it occurs, is generally accompanied by either anger or delight of the speaker at the dismay of the PTO. Thus, in most cases, the speaker of sar- castic comments were either angry with the PTO or rejoicing at its expense. Annotators were asked to differentiate between these two instances clearly using the context provided,

5

**450**

**450**

**450**

**451**

**451**

**451**

**452**

**452**

**452**

**453**

**453**

**453**

**454**

**454**

**454**

**455**

**455**

**455**

**456**

**456**

**456**

**457**

**457**

**457**

**458**

**458**

**458**

**459**

**459**

**459**

**460**

**460**

**460**

**461**

**461**

**461**

**462**

**462**

**462**

**463**

**463**

**463**

**464**

**464**

**464**

**465**

**465**

**465**

**466**

**466**

**466**

**467**

**467**

**467**

**468**

**468**

**468**

**469**

**469**

**469**

**470**

**470**

**470**

**471**

**471**

**471**

**472**

**472**

**472**

**473**

**473**

**473**

**474**

**474**

**474**

**475**

**475**

**475**

**476**

**476**

**476**

**477**

**477**

**477**

**478**

**478**

**478**

**479**

**479**

**479**

**480**

**480**

**480**

**481**

**481**

**481**

**482**

**482**

**482**

**483**

**483**

**483**

**484**

**484**

**484**

**485**

**485**

**485**

**486**

**486**

**486**

**487**

**487**

**487**

**488**

**488**

**488**

**489**

**489**

**489**

**490**

**490**

**490**

**491**

**491**

**491**

**492**

**492**

**492**

**493**

**493**

**493**

**494**

**494**

**494**

**495**

**495**

**495**

**496**

**496**

**496**

**497**

**497**

**497**

**498**

**498**

**498**

**499**

**499**

**499**

**500**

**501**

**502**

**503**

**504**

**505**

**506**

**507**

**508**

**509**

**510**

**511**

**512**

**513**

**514**

**515**

**516**

**517**

**518**

**519**

**520**

**521**

**522**

**523**

**524**

**525**

**526**

**527**

**528**

**529**

**530**

**531**

**532**

**533**

**534**

**535**

**536**

**537**

**538**

**539**

**540**

**541**

**542**

**543**

**544**

**545**

**546**

**547**

**548**

**549**

**NAACL-HLT 2019 Submission \*\*\*. Confidential Review Copy. DO NOT DISTRIBUTE.**

which was sometimes challenging. An exam- ple of sarcasm is presented in Example 4, in which the actual emotion expressed is *anger*, when it could be easily misunderstood to be *joy*. **.** *Example 4 -* हा हा हा! अब तम बताओग हम या बोल ? *(Ha Ha Ha ! Now you would tell me what I should speak?)* **Annotating Suspense** - Suspense was the toughest category for the annotators. Some- times, it proved very difficult for the anno- tators to know exactly when a sentence is of the category *suspense*. The annotators were asked to mark a sentence as suspense when there is some element in it which evokes a sense of anticipation or worry. Suspense is a unique feature of stories which does not get fully expressed in other types of written ma- terials such as news articles, formal reports, etc. Examples of such sentences are given in Example 5. **.** *Example 5 -* पछल पहर को मह फल म स नाटा हो गया| -हा क आवाज ब द हो गय | लीला न सोचा, या लोग कह चल गए, या सो गय? एकाएक स नाटा य छा गया? *(Last afternoon, the si- lence was over the entire place. There were no voices around. The sounds of Hu-Ha completely stopped. Leela thought, did people go somewhere, or perhaps they slept? Why all of a sudden there is silence everywhere?)* Next, we present the experiments performed for training the baseline models.

**5 Baseline Models**

In this section, we describe the baseline models that we train for the task of identifying one of the emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*, from a given sentence taken from a Hindi story. Both classic machine learning and modern deep learning models are trained. We report their performances and analyze the re- sults. We extensively use Sklearn (Pedregosa et al., 2011) and Keras (Chollet et al., 2018) as our machine learning toolkits.

**5.1 Dataset** The BHAAV dataset was randomly shuffled and split into train and test datasets with a ratio of 10:1. The distribution of labels in the two datasets are shown in Table 3. The proportion of distribution of labels in the test dataset is kept similar to the training dataset. We train our models on the training dataset and test the final predictions on the test dataset. We do not create a separate vali- dation dataset. However, we do use validation

data extracted from the training data, when- ever necessary for tuning the hyperparameters of the models.

data extracted from the training data, when- ever necessary for tuning the hyperparameters of the models.

data extracted from the training data, when- ever necessary for tuning the hyperparameters of the models.

data extracted from the training data, when- ever necessary for tuning the hyperparameters of the models.

**5.2 Text Preprocessing**

**5.2 Text Preprocessing**

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

Before training the classification models one needs to preprocess the text and represent each sentence as a feature vector. We tokenize each sentence into words and remove punc- tuations. We do not remove the stopwords. Since we deal with Hindi, the standard word tokenizers that are suitable for English lan- guage could not be used. Therefore, we used the tokenizer shipped with Classical Language Toolkit4. Each sentence is vectorized after a feature extraction step for the classic machine learning models such as Support Vector Ma- chines. Unigrams, Bigrams and Trigrams were generated as features for each sentence and their TF-IDF (Aizawa, 2003) scores were con- sidered as the feature values.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

One of the key components of the input fed to the deep learning models are pre-trained word embeddings (Kusner et al., 2015), that are used for representing each word of the in- put sentences by a dense real valued vector. Since the dataset on which we train our mod- els is relatively small, we use the pretrained word embeddings in order to prevent overfit- ting. This practice is commonly known as transfer learning5. We choose the Fasttext6 word embeddings (Bojanowski et al., 2016), trained on the Hindi Wikipedia corpus. This was a natural choice due to its easy availabil- ity. Additionally, Fasttext is possibly a bet- ter choice than other popular word embedding methods as it is more suitable for representing words belonging to morphologically rich lan- guages such as Hindi as described in Section 3.

While training the deep learning models, each sentence in the training and test dataset is converted to a fixed size document of 126 words (maximum length of a sentence in the dataset). Padding7 is used for sentences of length lesser than 126 words. Each word is

While training the deep learning models, each sentence in the training and test dataset is converted to a fixed size document of 126 words (maximum length of a sentence in the dataset). Padding7 is used for sentences of length lesser than 126 words. Each word is

While training the deep learning models, each sentence in the training and test dataset is converted to a fixed size document of 126 words (maximum length of a sentence in the dataset). Padding7 is used for sentences of length lesser than 126 words. Each word is

While training the deep learning models, each sentence in the training and test dataset is converted to a fixed size document of 126 words (maximum length of a sentence in the dataset). Padding7 is used for sentences of length lesser than 126 words. Each word is

While training the deep learning models, each sentence in the training and test dataset is converted to a fixed size document of 126 words (maximum length of a sentence in the dataset). Padding7 is used for sentences of length lesser than 126 words. Each word is

While training the deep learning models, each sentence in the training and test dataset is converted to a fixed size document of 126 words (maximum length of a sentence in the dataset). Padding7 is used for sentences of length lesser than 126 words. Each word is

While training the deep learning models, each sentence in the training and test dataset is converted to a fixed size document of 126 words (maximum length of a sentence in the dataset). Padding7 is used for sentences of length lesser than 126 words. Each word is

4http://docs.cltk.org/en/latest/hindi.html 5ftp://ftp.cs.wisc.edu/machine-learning/shavlik- group/torrey.handbook09.pdf

4http://docs.cltk.org/en/latest/hindi.html 5ftp://ftp.cs.wisc.edu/machine-learning/shavlik- group/torrey.handbook09.pdf

4http://docs.cltk.org/en/latest/hindi.html 5ftp://ftp.cs.wisc.edu/machine-learning/shavlik- group/torrey.handbook09.pdf

4http://docs.cltk.org/en/latest/hindi.html 5ftp://ftp.cs.wisc.edu/machine-learning/shavlik- group/torrey.handbook09.pdf

6https://github.com/facebookresearch/ fastText/blob/master/pretrained-vectors.md

6https://github.com/facebookresearch/ fastText/blob/master/pretrained-vectors.md

6https://github.com/facebookresearch/ fastText/blob/master/pretrained-vectors.md

7https://keras.io/preprocessing/sequence/

7https://keras.io/preprocessing/sequence/

6

**550**

**550**

**550**

**551**

**551**

**551**

**552**

**552**

**552**

**553**

**553**

**554**

**554**

**554**

**555**

**555**

**555**

**556**

**556**

**556**

**557**

**557**

**557**

**558**

**558**

**558**

**559**

**559**

**559**

**560**

**560**

**560**

**561**

**561**

**561**

**562**

**562**

**562**

**563**

**563**

**563**

**564**

**564**

**564**

**565**

**565**

**565**

**566**

**566**

**566**

**567**

**567**

**567**

**568**

**568**

**568**

**569**

**569**

**569**

**570**

**570**

**570**

**570**

**571**

**571**

**571**

**572**

**572**

**572**

**573**

**573**

**573**

**574**

**574**

**574**

**575**

**575**

**575**

**576**

**576**

**576**

**577**

**577**

**577**

**578**

**578**

**578**

**579**

**579**

**579**

**580**

**580**

**580**

**581**

**581**

**581**

**582**

**582**

**582**

**583**

**583**

**583**

**584**

**584**

**584**

**585**

**585**

**585**

**586**

**586**

**586**

**587**

**587**

**587**

**588**

**588**

**588**

**589**

**589**

**589**

**590**

**590**

**590**

**591**

**591**

**591**

**592**

**592**

**592**

**593**

**593**

**593**

**594**

**594**

**594**

**595**

**595**

**595**

**596**

**596**

**596**

**597**

**597**

**597**

**597**

**598**

**598**

**598**

**599**

**599**

**599**

**600**

**601**

**602**

**603**

**604**

**605**

**606**

**607**

**608**

**609**

**610**

**611**

**612**

**613**

**614**

**615**

**616**

**617**

**618**

**619**

**620**

**621**

**622**

**623**

**624**

**625**

**626**

**627**

**628**

**629**

**630**

**631**

**632**

**633**

**634**

**635**

**636**

**637**

**638**

**639**

**640**

**641**

**642**

**643**

**644**

**645**

**646**

**647**

**648**

**649**

**NAACL-HLT 2019 Submission \*\*\*. Confidential Review Copy. DO NOT DISTRIBUTE.**

represented as a 300 dimensional vector by the word embedding model. All the words in the dataset are mapped to their correspond- ing word embedding vector. Whenever a word is not found in the vocabulary of the word embedding model we assign it a 300 dimen- sional zero vector. Each sentence is then repre- sented as a matrix of its constituent words and their corresponding embedding vector, which is then fed as an input to the deep learning algorithms.

**Hyperparameter Range No. of Filters for CNN** 100, 200, 300, 400

**Filter sizes for the CNN model** 1, 2, 3, 4, 5, 6 **Dense Output Layer Size** 100, 200, 300, 400

**Dropout Probability** 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

**Learning Rate** 0.0001, 0.001

**Batch Sizes** 8, 16, 32, 64, 128

**Epochs** 10, 50, 100, 150 **LSTM units** 8, 16, 32, 64, 128, 256

Table 4: Hyperparameter ranges used for random search during training deep learning models (CNN and Bidirectional LSTM).

**Method Macro Avg**

**Precision**

perparameters among the ones shown in Table 4, that best fitted a fixed randomly selected validation data comprising of 20% of the train- ing data. Only 100 iterations of random search was performed. Once the hyperparameter tun- ing was done the final model was trained on the entire training data using the selected hy- perparameters. Adam (Kingma and Ba, 2014) with two annealing restarts has been shown to work faster and perform better than SGD in other NLP tasks (Denkowski and Neubig, 2017). Therefore, we use the same as our optimization algorithm for the deep learning models. As the task is a multi-class classifi- cation problem, categorical cross entropy was used as the loss function, and the final layer of both the deep learning models consisted of a fully-connected dense neural network with the extracted features as the input and a soft- max output giving the prediction probability for each of the five emotion categories.

Among the classic machine learning tech-

**Macro Avg Recall**

niques, *Support Vector Machine* (SVM) with a linear kernel (Hsu et al., 2003), *Logistic Re- gression* (Yu et al., 2011) and *Random Forests* (Breiman, 2001) were trained. A shallow Con- volutional Neural Network with a single in- put channel similar to (Severyn and Moschitti, 2015), and Bidirectional Long Short Term Memory networks with an architecture sim- ilar to (Mahata et al., 2018), are the deep learning models that were trained. A random classifier that randomly generated predictions from a label distribution similar to that of the training dataset was also implemented. Table 5 summarizes the performances of the classi- fiers on the test dataset for the following met- rics - *macro average precision*, *macro aver- age recall*, *macro average F1-score*, and *accu- racy* (Sokolova and Lapalme, 2009). We chose macro-average measures as the data is imbal- anced and macro-averaging will assign equal weights to all the categories, which gives a bet- ter generic performance of any classifier.

**6 Discussion**

In order to analyze the possible features chosen by a machine learning classification algorithm for discriminating between different categories of emotions and to validate the ability of the BHAAV dataset in providing such features to **Macro F1 Avg**

**Accuracy**

**Logistic Regression** 0.58 0.62 0.58 0.62 **SVM** 0.48 0.52 0.49 0.52 **Random Forests** 0.44 0.59 0.45 0.59 **CNN** 0.50 0.55 0.51 0.55 **BLSTM** 0.43 0.60 0.47 0.60 **Random Classifier** 0.40 0.40 0.40 0.40

Table 5: Performance of the baseline supervised clas- sification models on BHAAV dataset.

**Emotion Top 10 Important Unigram Features**

**joy**

स न (glad), सदर (beautiful), खश (happy), हस, (laugh), सगीत (music), खलौन (toys), मजा (fun), आनद (joy), हसकर (smilingly), उछल (jump)

**anger**

अपमान (insult), ग सा (anger), ोध (anger), बदला (revenge), मख (idiot), सजा (punishment), जह नम (hell), आग (fire), (evil), च लाया (screamed)

**suspense**

आवाज़ (sound), आ य (astonishment), ज न (Genie), दखा (saw), य (war), छनl (sound of anklets), कहा (where), जा (magic), अचानक (suddenly), जहाज (ship),

**sad**

रो (cry), मर (die), रोन (crying), ख (sadness), दय (heart), खी (sad), जीवन (life), आस (tears), रोत (cry), भगवान (God)

**neutral**

कसान (farmer), उसन (he), ब नी (Binny), पछा (asked), दादाजी (grandfather), कल (tomorrow), प डत (pundit), महता (mehta), मा (mother), आना (come)

Table 6: Top 10 most important features for each emo- tion category as identified by the Logistic Regression model during training.

**5.3 Training** All the machine learning models were trained after selecting the hyperparameters on a vali- dation data. 10-fold Cross Validation was used for the classic techniques. For the deep learn- ing models, random search (Bergstra and Ben- gio, 2012) was used for selecting the best hy-

7

**650**

**651**

**652**

**653**

**654**

**655**

**656**

**657**

**658**

**659**

**660**

**661**

**662**

**663**

**664**

**665**

**666**

**667**

**668**

**669**

**670**

**671**

**672**

**673**

**674**

**675**

**676**

**677**

**678**

**679**

**680**

**681**

**682**

**683**

**684**

**685**

**686**

**687**

**688**

**689**

**690**

**691**

**692**

**693**

**694**

**695**

**696**

**697**

**698**

**699**

**700**

**701**

**702**

**703**

**704**

**705**

**706**

**707**

**708**

**709**

**710**

**711**

**712**

**713**

**714**

**715**

**716**

**717**

**718**

**719**

**720**

**721**

**722**

**723**

**724**

**725**

**726**

**727**

**728**

**729**

**730**

**731**

**732**

**733**

**734**

**735**

**736**

**737**

**738**

**739**

**740**

**741**

**742**

**743**

**744**

**745**

**746**

**747**

**748**

**749**

(b) Exploiter and Exploited

Figure 1: Flow of emotions in randomly selected stories from two different genres.

any classifier, we looked at the most impor- tant features chosen by the Logistic Regres- sion model. Table 6 shows the top 10 most in- formative unigram features for each category of emotion chosen by the model in order to make the final predictions. As evident from the choices, words like स न (glad), सदर (beau- tiful), खश (happy), हस, (laugh), are sensible indicators of *joy*, and so are the words like अपमान (insult), ग सा (anger), ोध (anger), बदला (revenge), for *anger*. The other categories also show a similar pattern.

We also looked at the performance of the classifiers for individual categories. The cate- gory of *neutral* had the best performance con- sistently, which is quiet easy to guess from the the data distribution (Table 3) and it being the majority class. The performance of the *suspense* category was consistently low. Al- though, the category of *anger* had a similar presence in the dataset, yet it had better per- formance than *suspense*. This might be due to the presence of better discriminative fea- tures for *anger* than *suspense*. The other rea- son could be related to challenges associated with annotating the *suspense* category (Sec- tion 4.1).

Our analysis provides a brief insight into the BHAAV dataset from which we can con- clude that it is an appropriate dataset for emo- tion identification and classification tasks. Al- though, the dataset is created from stories, it can possibly be used for many other domains as it is rich in features indicating the five dif- ferent emotions as presented in this work. The annotations were done from the perspective of a reader/narrator trying to express the emo- tion of a sentence, given the existing scenario in the story and whenever applicable trying to express the emotion of a character in the story.

This also makes this dataset suitable for train- ing automated text-to-speech interfaces (e.g., audio books) for story narration and improv- ing them by infusing emotions in them.

This also makes this dataset suitable for train- ing automated text-to-speech interfaces (e.g., audio books) for story narration and improv- ing them by infusing emotions in them.

This also makes this dataset suitable for train- ing automated text-to-speech interfaces (e.g., audio books) for story narration and improv- ing them by infusing emotions in them.

This also makes this dataset suitable for train- ing automated text-to-speech interfaces (e.g., audio books) for story narration and improv- ing them by infusing emotions in them.

This also makes this dataset suitable for train- ing automated text-to-speech interfaces (e.g., audio books) for story narration and improv- ing them by infusing emotions in them.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

The dataset is also appropriate for analyz- ing the flow of emotions in individual stories and study them for different genres. We plot- ted the flow of emotions in a randomly picked story from two different genres as shown in Figure 1. It is observable from the figures that each story has its own distinct emotion foot- print. It would be interesting to study them and draw interesting linguistic insights from the Hindi literature. BHAAV can facilitate such experiments.

**7 Future Work and Conclusion**

**7 Future Work and Conclusion**

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

In this work we publicly shared the first and the largest annotated corpus - BHAAV with 20,304 sentences in Hindi, categorized as ex- pressing one of the five different emotions - *anger*, *joy*, *suspense*, *sad*, and *neutral*. The sentences are collected from 230 different short stories spanning across 10 genres. We pro- vided a detailed description of the dataset, lan- guage specific challenges, annotation process, challenges associated with annotations and re- ported performances of the baseline classifica- tion models trained on the dataset for identify- ing emotions expressed in a sentence. Through different observations we confirm the dataset to be rich with emotion cues and point to the potential applications of the dataset. In the future, we plan to work on enriching the dataset with more annotations related to sen- timent and discourse analysis. We believe that BHAAV will prove to be a valuable resource in Hindi and encourage further experiments in the domain of emotion analysis from text.

**NAACL-HLT 2019 Submission \*\*\*. Confidential Review Copy. DO NOT DISTRIBUTE.**

(a) Idealist

(a) Idealist

**750**

**751**

**752**

**753**

**754**

**755**

**756**

**756**

**756**

**757**

**758**

**758**

**759**

**759**

**760**

**761**

**761**

**761**

**762**

**762**

**762**

**763**

**763**

**763**

**764**

**764**

**764**

**765**

**765**

**765**

**766**

**766**

**766**

**767**

**767**

**767**

**768**

**768**

**768**

**769**

**769**

**769**

**770**

**770**

**770**

**771**

**771**

**771**

**772**

**772**

**772**

**773**

**773**

**773**

**774**

**774**

**774**

**775**

**775**

**775**

**776**

**776**

**776**

**777**

**777**

**778**

**778**

**778**

**779**

**779**

**779**

**780**

**780**

**780**

**781**

**781**

**781**

**782**

**782**

**782**

**783**

**783**

**783**

**784**

**784**

**784**

**785**

**785**

**785**

**786**

**786**

**786**

**787**

**787**

**787**

**788**

**788**

**788**

**789**

**789**

**789**

**790**

**790**

**790**

**791**

**791**

**791**

**792**

**792**

**792**

**793**

**793**

**793**

**794**

**794**

**794**

**795**

**795**

**795**

**796**

**796**

**796**

**797**

**797**

**797**

**798**

**798**

**798**

**799**

**799**

8

**800**

**801**

**802**

**803**

**804**

**805**

**806**

**807**

**808**

**809**

**810**

**811**

**812**

**813**

**814**

**815**

**816**

**817**

**818**

**819**

**820**

**821**

**822**

**823**

**824**

**825**

**826**

**827**

**828**

**829**

**830**

**831**

**832**

**833**

**834**

**835**

**836**

**837**

**838**

**839**

**840**

**841**

**842**

**843**

**844**

**845**

**846**

**847**

**848**

**849**

**NAACL-HLT 2019 Submission \*\*\*. Confidential Review Copy. DO NOT DISTRIBUTE.**

**References**

Akiko Aizawa. 2003. An information-theoretic per- spective of tf–idf measures. *Information Pro-*

*cessing & Management*, 39(1):45–65.

Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of the conference on human lan- guage technology and empirical methods in nat- ural language processing*, pages 579–586. Asso- ciation for Computational Linguistics.

Piyush Arora. 2013. Sentiment analysis for hindi language. *MS by Research in Computer Science*.

AR Balamurali, Aditya Joshi, and Pushpak Bhat- tacharyya. 2012. Cross-lingual sentiment anal- ysis for indian languages using linked wordnets. *Proceedings of COLING 2012: Posters*, pages 73–82.

James Bergstra and Yoshua Bengio. 2012. Random search for hyper-parameter optimization. *Jour- nal of Machine Learning Research*, 13(Feb):281– 305.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching word vec- tors with subword information. *arXiv preprint arXiv:1607.04606*.

Roger Bougie, Rik Pieters, and Marcel Zeelenberg. 2003. Angry customers don’t come back, they get back: The experience and behavioral impli- cations of anger and dissatisfaction in services. *Journal of the Academy of Marketing Science*, 31(4):377–393.

Cynthia Breazeal and Rodney Brooks. 2005. Robot emotion: A functional perspective. *Who*

*needs emotions*, pages 271–310.

Leo Breiman. 2001. Random forests. *Machine*

*learning*, 45(1):5–32.

Erik Cambria, Soujanya Poria, Alexander Gel- bukh, and Mike Thelwall. 2017. Sentiment anal- ysis is a big suitcase. *IEEE Intelligent Systems*, 32(6):74–80.

François Chollet et al. 2018. Keras: The python deep learning library. *Astrophysics Source Code*

*Library*.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting de- pression via social media. *ICWSM*, 13:1–10.

Michael Denkowski and Graham Neubig. 2017. Stronger baselines for trustable results in neural machine translation. *arXiv preprint arXiv:1706.09733*.

Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intra- class correlation coefficient as measures of reli- ability. *Educational and psychological measure- ment*, 33(3):613–619.

Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intra- class correlation coefficient as measures of reli- ability. *Educational and psychological measure- ment*, 33(3):613–619.

Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intra- class correlation coefficient as measures of reli- ability. *Educational and psychological measure- ment*, 33(3):613–619.

Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intra- class correlation coefficient as measures of reli- ability. *Educational and psychological measure- ment*, 33(3):613–619.

Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intra- class correlation coefficient as measures of reli- ability. *Educational and psychological measure- ment*, 33(3):613–619.

Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intra- class correlation coefficient as measures of reli- ability. *Educational and psychological measure- ment*, 33(3):613–619.

Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intra- class correlation coefficient as measures of reli- ability. *Educational and psychological measure- ment*, 33(3):613–619.

Virginia Francisco and Pablo Gervás. 2006. Au- tomated mark up of affective information in en- glish texts. In *International Conference on Text, Speech and Dialogue*, pages 375–382. Springer.

Virginia Francisco and Pablo Gervás. 2006. Au- tomated mark up of affective information in en- glish texts. In *International Conference on Text, Speech and Dialogue*, pages 375–382. Springer.

Virginia Francisco and Pablo Gervás. 2006. Au- tomated mark up of affective information in en- glish texts. In *International Conference on Text, Speech and Dialogue*, pages 375–382. Springer.

Virginia Francisco and Pablo Gervás. 2006. Au- tomated mark up of affective information in en- glish texts. In *International Conference on Text, Speech and Dialogue*, pages 375–382. Springer.

Kai Gao, Hua Xu, and Jiushuo Wang. 2015. A rule-based approach to emotion cause detection for chinese micro-blogs. *Expert Systems with Applications*, 42(9):4517–4528.

Kai Gao, Hua Xu, and Jiushuo Wang. 2015. A rule-based approach to emotion cause detection for chinese micro-blogs. *Expert Systems with Applications*, 42(9):4517–4528.

Kai Gao, Hua Xu, and Jiushuo Wang. 2015. A rule-based approach to emotion cause detection for chinese micro-blogs. *Expert Systems with Applications*, 42(9):4517–4528.

Kai Gao, Hua Xu, and Jiushuo Wang. 2015. A rule-based approach to emotion cause detection for chinese micro-blogs. *Expert Systems with Applications*, 42(9):4517–4528.

Narendra Gupta, Mazin Gilbert, and Giuseppe Di Fabbrizio. 2013. Emotion detection in email customer care. *Computational Intelligence*, 29(3):489–505.

Narendra Gupta, Mazin Gilbert, and Giuseppe Di Fabbrizio. 2013. Emotion detection in email customer care. *Computational Intelligence*, 29(3):489–505.

Narendra Gupta, Mazin Gilbert, and Giuseppe Di Fabbrizio. 2013. Emotion detection in email customer care. *Computational Intelligence*, 29(3):489–505.

Narendra Gupta, Mazin Gilbert, and Giuseppe Di Fabbrizio. 2013. Emotion detection in email customer care. *Computational Intelligence*, 29(3):489–505.

Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. 2003. A practical guide to support vector classification.

Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. 2003. A practical guide to support vector classification.

Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. 2003. A practical guide to support vector classification.

Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. 2003. A practical guide to support vector classification.

Aditya Joshi, AR Balamurali, and Pushpak Bhat- tacharyya. 2010. A fall-back strategy for senti- ment analysis in hindi: a case study. *Proceed- ings of the 8th ICON*.

Aditya Joshi, AR Balamurali, and Pushpak Bhat- tacharyya. 2010. A fall-back strategy for senti- ment analysis in hindi: a case study. *Proceed- ings of the 8th ICON*.

Aditya Joshi, AR Balamurali, and Pushpak Bhat- tacharyya. 2010. A fall-back strategy for senti- ment analysis in hindi: a case study. *Proceed- ings of the 8th ICON*.

Aditya Joshi, AR Balamurali, and Pushpak Bhat- tacharyya. 2010. A fall-back strategy for senti- ment analysis in hindi: a case study. *Proceed- ings of the 8th ICON*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kathrin Knautz, Tobias Siebenlist, and Wolf- gang G Stock. 2010. Memose: search engine for emotions in multimedia documents. In *Pro- ceedings of the 33rd International ACM SIGIR Conference on Research and development in in- formation retrieval*, pages 791–792. ACM.

Kathrin Knautz, Tobias Siebenlist, and Wolf- gang G Stock. 2010. Memose: search engine for emotions in multimedia documents. In *Pro- ceedings of the 33rd International ACM SIGIR Conference on Research and development in in- formation retrieval*, pages 791–792. ACM.

Kathrin Knautz, Tobias Siebenlist, and Wolf- gang G Stock. 2010. Memose: search engine for emotions in multimedia documents. In *Pro- ceedings of the 33rd International ACM SIGIR Conference on Research and development in in- formation retrieval*, pages 791–792. ACM.

Kathrin Knautz, Tobias Siebenlist, and Wolf- gang G Stock. 2010. Memose: search engine for emotions in multimedia documents. In *Pro- ceedings of the 33rd International ACM SIGIR Conference on Research and development in in- formation retrieval*, pages 791–792. ACM.

Kathrin Knautz, Tobias Siebenlist, and Wolf- gang G Stock. 2010. Memose: search engine for emotions in multimedia documents. In *Pro- ceedings of the 33rd International ACM SIGIR Conference on Research and development in in- formation retrieval*, pages 791–792. ACM.

Kathrin Knautz, Tobias Siebenlist, and Wolf- gang G Stock. 2010. Memose: search engine for emotions in multimedia documents. In *Pro- ceedings of the 33rd International ACM SIGIR Conference on Research and development in in- formation retrieval*, pages 791–792. ACM.

Kathrin Knautz, Tobias Siebenlist, and Wolf- gang G Stock. 2010. Memose: search engine for emotions in multimedia documents. In *Pro- ceedings of the 33rd International ACM SIGIR Conference on Research and development in in- formation retrieval*, pages 791–792. ACM.

Shashidhar G Koolagudi, Ramu Reddy, Jainath Yadav, and K Sreenivasa Rao. 2011. Iitkgp- sehsc: Hindi speech corpus for emotion anal- ysis. In *Devices and Communications (ICDe- Com), 2011 International Conference on*, pages 1–5. IEEE.

Shashidhar G Koolagudi, Ramu Reddy, Jainath Yadav, and K Sreenivasa Rao. 2011. Iitkgp- sehsc: Hindi speech corpus for emotion anal- ysis. In *Devices and Communications (ICDe- Com), 2011 International Conference on*, pages 1–5. IEEE.

Shashidhar G Koolagudi, Ramu Reddy, Jainath Yadav, and K Sreenivasa Rao. 2011. Iitkgp- sehsc: Hindi speech corpus for emotion anal- ysis. In *Devices and Communications (ICDe- Com), 2011 International Conference on*, pages 1–5. IEEE.

Shashidhar G Koolagudi, Ramu Reddy, Jainath Yadav, and K Sreenivasa Rao. 2011. Iitkgp- sehsc: Hindi speech corpus for emotion anal- ysis. In *Devices and Communications (ICDe- Com), 2011 International Conference on*, pages 1–5. IEEE.

Shashidhar G Koolagudi, Ramu Reddy, Jainath Yadav, and K Sreenivasa Rao. 2011. Iitkgp- sehsc: Hindi speech corpus for emotion anal- ysis. In *Devices and Communications (ICDe- Com), 2011 International Conference on*, pages 1–5. IEEE.

Shashidhar G Koolagudi, Ramu Reddy, Jainath Yadav, and K Sreenivasa Rao. 2011. Iitkgp- sehsc: Hindi speech corpus for emotion anal- ysis. In *Devices and Communications (ICDe- Com), 2011 International Conference on*, pages 1–5. IEEE.

Shashidhar G Koolagudi, Ramu Reddy, Jainath Yadav, and K Sreenivasa Rao. 2011. Iitkgp- sehsc: Hindi speech corpus for emotion anal- ysis. In *Devices and Communications (ICDe- Com), 2011 International Conference on*, pages 1–5. IEEE.

Shashidhar G Koolagudi, Ramu Reddy, Jainath Yadav, and K Sreenivasa Rao. 2011. Iitkgp- sehsc: Hindi speech corpus for emotion anal- ysis. In *Devices and Communications (ICDe- Com), 2011 International Conference on*, pages 1–5. IEEE.

Shashidhar G Koolagudi, Ramu Reddy, Jainath Yadav, and K Sreenivasa Rao. 2011. Iitkgp- sehsc: Hindi speech corpus for emotion anal- ysis. In *Devices and Communications (ICDe- Com), 2011 International Conference on*, pages 1–5. IEEE.

Shashidhar G Koolagudi, Ramu Reddy, Jainath Yadav, and K Sreenivasa Rao. 2011. Iitkgp- sehsc: Hindi speech corpus for emotion anal- ysis. In *Devices and Communications (ICDe- Com), 2011 International Conference on*, pages 1–5. IEEE.

Zoltán Kövecses. 2003. *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge University Press.

Zoltán Kövecses. 2003. *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge University Press.

Zoltán Kövecses. 2003. *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge University Press.

Zoltán Kövecses. 2003. *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge University Press.

Klaus Krippendorff. 2011. Computing krippen-

Klaus Krippendorff. 2011. Computing krippen-

dorff’s alpha-reliability.

dorff’s alpha-reliability.

dorff’s alpha-reliability.

Matt Kusner, Yu Sun, Nicholas Kolkin, and Kil- ian Weinberger. 2015. From word embeddings to document distances. In *International Con- ference on Machine Learning*, pages 957–966.

Matt Kusner, Yu Sun, Nicholas Kolkin, and Kil- ian Weinberger. 2015. From word embeddings to document distances. In *International Con- ference on Machine Learning*, pages 957–966.

Matt Kusner, Yu Sun, Nicholas Kolkin, and Kil- ian Weinberger. 2015. From word embeddings to document distances. In *International Con- ference on Machine Learning*, pages 957–966.

Matt Kusner, Yu Sun, Nicholas Kolkin, and Kil- ian Weinberger. 2015. From word embeddings to document distances. In *International Con- ference on Machine Learning*, pages 957–966.

Matt Kusner, Yu Sun, Nicholas Kolkin, and Kil- ian Weinberger. 2015. From word embeddings to document distances. In *International Con- ference on Machine Learning*, pages 957–966.

Zhang Lei, Wang Shuai, and Liu Bing. 2018. Deep learning for sentiment analysis: A survey. *Cor- nell Science Library*.

Zhang Lei, Wang Shuai, and Liu Bing. 2018. Deep learning for sentiment analysis: A survey. *Cor- nell Science Library*.

Zhang Lei, Wang Shuai, and Liu Bing. 2018. Deep learning for sentiment analysis: A survey. *Cor- nell Science Library*.

Zhang Lei, Wang Shuai, and Liu Bing. 2018. Deep learning for sentiment analysis: A survey. *Cor- nell Science Library*.

9

**850**

**850**

**850**

**851**

**851**

**851**

**852**

**852**

**852**

**853**

**853**

**853**

**854**

**854**

**855**

**855**

**855**

**856**

**856**

**856**

**857**

**857**

**857**

**858**

**858**

**859**

**859**

**859**

**860**

**860**

**861**

**861**

**861**

**862**

**862**

**863**

**863**

**863**

**864**

**864**

**864**

**865**

**865**

**865**

**866**

**866**

**867**

**867**

**868**

**868**

**868**

**869**

**869**

**870**

**870**

**870**

**871**

**871**

**872**

**872**

**872**

**873**

**873**

**874**

**874**

**874**

**875**

**875**

**875**

**876**

**876**

**877**

**877**

**877**

**878**

**878**

**878**

**879**

**879**

**879**

**880**

**880**

**880**

**881**

**881**

**881**

**882**

**882**

**882**

**882**

**883**

**883**

**883**

**884**

**884**

**884**

**885**

**885**

**885**

**886**

**886**

**886**

**887**

**887**

**888**

**888**

**888**

**889**

**889**

**889**

**890**

**890**

**890**

**891**

**891**

**891**

**892**

**892**

**892**

**893**

**893**

**893**

**894**

**894**

**894**

**895**

**895**

**895**

**896**

**896**

**896**

**897**

**897**

**897**

**898**

**898**

**898**

**899**

**899**

**899**

**900**

**901**

**902**

**903**

**904**

**905**

**906**

**907**

**908**

**909**

**910**

**911**

**912**

**913**

**914**

**915**

**916**

**917**

**918**

**919**

**920**

**921**

**922**

**923**

**924**

**925**

**926**

**927**

**928**

**929**

**930**

**931**

**932**

**933**

**934**

**935**

**936**

**937**

**938**

**939**

**940**

**941**

**942**

**943**

**944**

**945**

**946**

**947**

**948**

**949**

**NAACL-HLT 2019 Submission \*\*\*. Confidential Review Copy. DO NOT DISTRIBUTE.**

Diane J Litman and Kate Forbes-Riley. 2004. Pre- dicting student emotions in computer-human tu- toring dialogues. In *Proceedings of the 42nd An- nual Meeting on Association for Computational Linguistics*, page 351. Association for Computa- tional Linguistics.

Debanjan Mahata, Jasper Friedrichs, Rajiv Ratn Shah, et al. 2018. # phramacovigilance- exploring deep learning techniques for identify- ing mentions of medication intake from twitter. *arXiv preprint arXiv:1805.06375*.

Namita Mittal, Basant Agarwal, Garvit Chouhan, Nitin Bania, and Prateek Pareek. 2013. Senti- ment analysis of hindi reviews based on negation and discourse relation. In *Proceedings of the 11th Workshop on Asian Language Resources*, pages 45–50.

Saif Mohammad. 2016. A practical guide to senti- ment annotation: Challenges and solutions. In *Proceedings of the 7th Workshop on Computa- tional Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 174–179.

Saif M Mohammad. 2017. Challenges in sentiment analysis. In *A Practical Guide to Sentiment*

*Analysis*, pages 61–83. Springer.

Saif M Mohammad and Peter D Turney. 2013. Nrc emotion lexicon. *National Research Coun-*

*cil, Canada*.

Andrew Ortony, Gerald L Clore, and Mark A Foss. 1987. The referential structure of the affective

lexicon. *Cognitive science*, 11(3):341–364.

Bo Pang, Lillian Lee, et al. 2008. Opinion min- ing and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135.

Braja Gopal Patra, Dipankar Das, Amitava Das, and Rajendra Prasath. 2015. Shared task on sentiment analysis in indian languages (sail) tweets-an overview. In *International Conference on Mining Intelligence and Knowledge Explo- ration*, pages 650–655. Springer.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Pretten- hofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in python. *Jour- nal of machine learning research*, 12(Oct):2825– 2830.

James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erl- baum Associates*, 71(2001):2001.

Robert Plutchik. 1984. Emotions: A general psy- choevolutionary theory. *Approaches to emotion*, 1984:197–219.

Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective com- puting: From unimodal analysis to multimodal fusion. *Information Fusion*, 37:98–125.

Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective com- puting: From unimodal analysis to multimodal fusion. *Information Fusion*, 37:98–125.

Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective com- puting: From unimodal analysis to multimodal fusion. *Information Fusion*, 37:98–125.

Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective com- puting: From unimodal analysis to multimodal fusion. *Information Fusion*, 37:98–125.

Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective com- puting: From unimodal analysis to multimodal fusion. *Information Fusion*, 37:98–125.

Niklas Ravaja, Timo Saari, Marko Turpeinen, Jari Laarni, Mikko Salminen, and Matias Kivikan- gas. 2006. Spatial presence and emotions during video game playing: Does it matter with whom you play? *Presence: Teleoperators and Virtual Environments*, 15(4):381–392.

Niklas Ravaja, Timo Saari, Marko Turpeinen, Jari Laarni, Mikko Salminen, and Matias Kivikan- gas. 2006. Spatial presence and emotions during video game playing: Does it matter with whom you play? *Presence: Teleoperators and Virtual Environments*, 15(4):381–392.

Niklas Ravaja, Timo Saari, Marko Turpeinen, Jari Laarni, Mikko Salminen, and Matias Kivikan- gas. 2006. Spatial presence and emotions during video game playing: Does it matter with whom you play? *Presence: Teleoperators and Virtual Environments*, 15(4):381–392.

Niklas Ravaja, Timo Saari, Marko Turpeinen, Jari Laarni, Mikko Salminen, and Matias Kivikan- gas. 2006. Spatial presence and emotions during video game playing: Does it matter with whom you play? *Presence: Teleoperators and Virtual Environments*, 15(4):381–392.

Niklas Ravaja, Timo Saari, Marko Turpeinen, Jari Laarni, Mikko Salminen, and Matias Kivikan- gas. 2006. Spatial presence and emotions during video game playing: Does it matter with whom you play? *Presence: Teleoperators and Virtual Environments*, 15(4):381–392.

Niklas Ravaja, Timo Saari, Marko Turpeinen, Jari Laarni, Mikko Salminen, and Matias Kivikan- gas. 2006. Spatial presence and emotions during video game playing: Does it matter with whom you play? *Presence: Teleoperators and Virtual Environments*, 15(4):381–392.

Niklas Ravaja, Timo Saari, Marko Turpeinen, Jari Laarni, Mikko Salminen, and Matias Kivikan- gas. 2006. Spatial presence and emotions during video game playing: Does it matter with whom you play? *Presence: Teleoperators and Virtual Environments*, 15(4):381–392.

Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Sci- ence*, 5(1):31.

Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Sci- ence*, 5(1):31.

Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Sci- ence*, 5(1):31.

Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Sci- ence*, 5(1):31.

Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Sci- ence*, 5(1):31.

Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Sci- ence*, 5(1):31.

Monika Schwarz-Friesel. 2015. Language and emo- tion. *The Cognitive Linguistic Perspective, in: Ulrike Lüdtke (Hg.), Emotion in Lan- guage. Theory–Research–Application, Amster- dam*, pages 157–173.

Monika Schwarz-Friesel. 2015. Language and emo- tion. *The Cognitive Linguistic Perspective, in: Ulrike Lüdtke (Hg.), Emotion in Lan- guage. Theory–Research–Application, Amster- dam*, pages 157–173.

Monika Schwarz-Friesel. 2015. Language and emo- tion. *The Cognitive Linguistic Perspective, in: Ulrike Lüdtke (Hg.), Emotion in Lan- guage. Theory–Research–Application, Amster- dam*, pages 157–173.

Monika Schwarz-Friesel. 2015. Language and emo- tion. *The Cognitive Linguistic Perspective, in: Ulrike Lüdtke (Hg.), Emotion in Lan- guage. Theory–Research–Application, Amster- dam*, pages 157–173.

Monika Schwarz-Friesel. 2015. Language and emo- tion. *The Cognitive Linguistic Perspective, in: Ulrike Lüdtke (Hg.), Emotion in Lan- guage. Theory–Research–Application, Amster- dam*, pages 157–173.

Monika Schwarz-Friesel. 2015. Language and emo- tion. *The Cognitive Linguistic Perspective, in: Ulrike Lüdtke (Hg.), Emotion in Lan- guage. Theory–Research–Application, Amster- dam*, pages 157–173.

Monika Schwarz-Friesel. 2015. Language and emo- tion. *The Cognitive Linguistic Perspective, in: Ulrike Lüdtke (Hg.), Emotion in Lan- guage. Theory–Research–Application, Amster- dam*, pages 157–173.

Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter sentiment analysis with deep convolu- tional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Re- trieval*, pages 959–962. ACM.

Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter sentiment analysis with deep convolu- tional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Re- trieval*, pages 959–962. ACM.

Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter sentiment analysis with deep convolu- tional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Re- trieval*, pages 959–962. ACM.

Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter sentiment analysis with deep convolu- tional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Re- trieval*, pages 959–962. ACM.

Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter sentiment analysis with deep convolu- tional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Re- trieval*, pages 959–962. ACM.

Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter sentiment analysis with deep convolu- tional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Re- trieval*, pages 959–962. ACM.

Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter sentiment analysis with deep convolu- tional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Re- trieval*, pages 959–962. ACM.

Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter sentiment analysis with deep convolu- tional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Re- trieval*, pages 959–962. ACM.

Ameneh Gholipour Shahraki and Osmar R Zaiane. 2017. Lexical and learning-based emotion min- ing from text. In *Proceedings of the Interna- tional Conference on Computational Linguistics and Intelligent Text Processing*.

Ameneh Gholipour Shahraki and Osmar R Zaiane. 2017. Lexical and learning-based emotion min- ing from text. In *Proceedings of the Interna- tional Conference on Computational Linguistics and Intelligent Text Processing*.

Ameneh Gholipour Shahraki and Osmar R Zaiane. 2017. Lexical and learning-based emotion min- ing from text. In *Proceedings of the Interna- tional Conference on Computational Linguistics and Intelligent Text Processing*.

Ameneh Gholipour Shahraki and Osmar R Zaiane. 2017. Lexical and learning-based emotion min- ing from text. In *Proceedings of the Interna- tional Conference on Computational Linguistics and Intelligent Text Processing*.

Ameneh Gholipour Shahraki and Osmar R Zaiane. 2017. Lexical and learning-based emotion min- ing from text. In *Proceedings of the Interna- tional Conference on Computational Linguistics and Intelligent Text Processing*.

Ameneh Gholipour Shahraki and Osmar R Zaiane. 2017. Lexical and learning-based emotion min- ing from text. In *Proceedings of the Interna- tional Conference on Computational Linguistics and Intelligent Text Processing*.

Ameneh Gholipour Shahraki and Osmar R Zaiane. 2017. Lexical and learning-based emotion min- ing from text. In *Proceedings of the Interna- tional Conference on Computational Linguistics and Intelligent Text Processing*.

Marina Sokolova and Guy Lapalme. 2009. A sys- tematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437.

Marina Sokolova and Guy Lapalme. 2009. A sys- tematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437.

Marina Sokolova and Guy Lapalme. 2009. A sys- tematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437.

Marina Sokolova and Guy Lapalme. 2009. A sys- tematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437.

Marina Sokolova and Guy Lapalme. 2009. A sys- tematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437.

Marina Sokolova and Guy Lapalme. 2009. A sys- tematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437.

Marina Sokolova and Guy Lapalme. 2009. A sys- tematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Pro- ceedings of the 4th international workshop on semantic evaluations*, pages 70–74. Association for Computational Linguistics.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Pro- ceedings of the 4th international workshop on semantic evaluations*, pages 70–74. Association for Computational Linguistics.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Pro- ceedings of the 4th international workshop on semantic evaluations*, pages 70–74. Association for Computational Linguistics.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Pro- ceedings of the 4th international workshop on semantic evaluations*, pages 70–74. Association for Computational Linguistics.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Pro- ceedings of the 4th international workshop on semantic evaluations*, pages 70–74. Association for Computational Linguistics.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Pro- ceedings of the 4th international workshop on semantic evaluations*, pages 70–74. Association for Computational Linguistics.

Carlo Strapparava, Alessandro Valitutti, et al. 2004. Wordnet affect: an affective extension of wordnet. In *Lrec*, volume 4, pages 1083–1086. Citeseer.

Carlo Strapparava, Alessandro Valitutti, et al. 2004. Wordnet affect: an affective extension of wordnet. In *Lrec*, volume 4, pages 1083–1086. Citeseer.

Carlo Strapparava, Alessandro Valitutti, et al. 2004. Wordnet affect: an affective extension of wordnet. In *Lrec*, volume 4, pages 1083–1086. Citeseer.

Carlo Strapparava, Alessandro Valitutti, et al. 2004. Wordnet affect: an affective extension of wordnet. In *Lrec*, volume 4, pages 1083–1086. Citeseer.

Carlo Strapparava, Alessandro Valitutti, et al. 2004. Wordnet affect: an affective extension of wordnet. In *Lrec*, volume 4, pages 1083–1086. Citeseer.

Carlo Strapparava, Alessandro Valitutti, et al. 2004. Wordnet affect: an affective extension of wordnet. In *Lrec*, volume 4, pages 1083–1086. Citeseer.

Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. Corpus creation and emotion prediction for hindi-english code-mixed social media text. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 128–135.

Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. Corpus creation and emotion prediction for hindi-english code-mixed social media text. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 128–135.

Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. Corpus creation and emotion prediction for hindi-english code-mixed social media text. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 128–135.

Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. Corpus creation and emotion prediction for hindi-english code-mixed social media text. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 128–135.

Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. Corpus creation and emotion prediction for hindi-english code-mixed social media text. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 128–135.

Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. Corpus creation and emotion prediction for hindi-english code-mixed social media text. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 128–135.

Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. Corpus creation and emotion prediction for hindi-english code-mixed social media text. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 128–135.

Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. Corpus creation and emotion prediction for hindi-english code-mixed social media text. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 128–135.

Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. Corpus creation and emotion prediction for hindi-english code-mixed social media text. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 128–135.

Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R Zaiane. 2017. Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, 50(2):25.

Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R Zaiane. 2017. Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, 50(2):25.

Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R Zaiane. 2017. Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, 50(2):25.

Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R Zaiane. 2017. Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, 50(2):25.

Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R Zaiane. 2017. Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, 50(2):25.

Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R Zaiane. 2017. Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, 50(2):25.

**950**

**950**

**950**

**951**

**951**

**951**

**952**

**952**

**952**

**953**

**953**

**953**

**954**

**954**

**954**

**955**

**955**

**955**

**956**

**956**

**956**

**957**

**957**

**957**

**958**

**958**

**958**

**959**

**959**

**959**

**960**

**960**

**960**

**961**

**961**

**961**

**962**

**962**

**962**

**963**

**963**

**963**

**964**

**964**

**964**

**965**

**965**

**965**

**966**

**966**

**966**

**967**

**967**

**967**

**968**

**968**

**968**

**969**

**969**

**969**

**970**

**970**

**970**

**971**

**971**

**971**

**971**

**972**

**972**

**972**

**973**

**973**

**973**

**974**

**974**

**974**

**974**

**975**

**975**

**975**

**976**

**976**

**976**

**977**

**977**

**977**

**977**

**978**

**978**

**978**

**979**

**979**

**979**

**980**

**980**

**980**

**981**

**981**

**981**

**982**

**982**

**982**

**983**

**983**

**983**

**984**

**984**

**984**

**985**

**985**

**985**

**986**

**986**

**986**

**987**

**987**

**987**

**988**

**988**

**988**

**989**

**989**

**989**

**990**

**990**

**990**

**991**

**991**

**991**

**992**

**992**

**992**

**993**

**993**

**993**

**994**

**994**

**994**

**995**

**995**

**995**

**996**

**996**

**996**

**997**

**997**

**997**

**998**

**998**

**998**

**999**

**999**

**999**

10

**1000**

**1001**

**1002**

**1003**

**1004**

**1005**

**1006**

**1007**

**1008**

**1009**

**1010**

**1011**

**1012**

**1013**

**1014**

**1015**

**1016**

**1017**

**1018**

**1019**

**1020**

**1021**

**1022**

**1023**

**1024**

**1025**

**1026**

**1027**

**1028**

**1029**

**1030**

**1031**

**1032**

**1033**

**1034**

**1035**

**1036**

**1037**

**1038**

**1039**

**1040**

**1041**

**1042**

**1043**

**1044**

**1045**

**1046**

**1047**

**1048**

**1049**

**NAACL-HLT 2019 Submission \*\*\*. Confidential Review Copy. DO NOT DISTRIBUTE.**

Changhua Yang, Kevin Hsin-Yih Lin, and Hsin- Hsi Chen. 2007. Emotion classification us- ing web blog corpora. In *Web Intelligence, IEEE/WIC/ACM International Conference on*, pages 275–278. IEEE.

Hsiang-Fu Yu, Fang-Lan Huang, and Chih-Jen Lin. 2011. Dual coordinate descent methods for lo- gistic regression and maximum entropy models. *Machine Learning*, 85(1-2):41–75.

11

**1050**

**1050**

**1051**

**1051**

**1052**

**1052**

**1053**

**1053**

**1054**

**1054**

**1055**

**1055**

**1056**

**1056**

**1057**

**1057**

**1058**

**1059**

**1060**

**1061**

**1062**

**1063**

**1064**

**1065**

**1066**

**1067**

**1068**

**1069**

**1070**

**1071**

**1072**

**1073**

**1074**

**1075**

**1076**

**1077**

**1078**

**1079**

**1080**

**1081**

**1082**

**1083**

**1084**

**1085**

**1086**

**1087**

**1088**

**1089**

**1090**

**1091**

**1092**

**1093**

**1094**

**1095**

**1096**

**1097**

**1098**

**1099**