

CHAPTER 1

INTRODUCTION

This chapter aims to make familiar with the necessity of this research and relevant concepts, the background behind this research, problems, and research questions associated with this field of interest. Besides, this chapter includes a comprehensive summary of research objectives, significance, and scope. Moreover, the expected contribution of this research is highlighted at the end of this chapter.

1.1 Introduction

The availability of information on the web increases day-by-day¹, especially in a real-time information network such as social networking websites (e.g., Facebook and Twitter, etc.), and those data and information (i.e., students' feedback) cannot be kept useless. Hence, to make them useful (i.e., to the information of decision making on students' satisfaction towards the services), there require analyzing those data and information, and this analysis is known as opinion mining (OM) or sentiment analysis (SA). Analyzing feedback of the students using SA techniques could either identify the students' emotions towards the current teaching is positive or negative (Altrabsheh et al., 2013). In general, sentiment analysis is the natural language processing (NLP) problem, and it is a field of computational study and classification process of opinions, attitudes, and emotions of people toward an entity about like/dislike (Medhat et al., 2014).

¹<https://datareportal.com/reports/digital-2020-global-digital-overview>

Sentiment analysis penetrates different areas of study such as NLP, text processing and analysis, computational linguistics, and biometrics. A good amount of research has been done on different types of data. Therefore, based on data, sentiment analysis could be classified as sentiment analysis from the text, emoticons, and multimode (image, audio, and video) (Yadav& Pandya, 2017; Soleymani et al., 2017; Kaur & Kautish, 2019). Besides, a new door is open for research in sentiment analysis, known as multilingual sentiment analysis (MLSA), where users (Example: Students, Teachers, etc. as per the data used in this research) could now post opinions in different human languages available in the world, such as Bengali, instead of only posting the text in English. These posts and comments may contain helpful information (sentiments) for the concern to whom it was posted or commented, and processing them is of great importance to those concerns (Hussain et al., 2016). Thus, transforming those comments and posts to bring out users' sentiments is known as multilingual sentiment analysis.

There is a good deal of research work done recently on sentiment analysis in a wide range of languages. However, due to the lack of proper sentiment dictionaries (knowledge bases or polarity lexicons) in languages other than English, those research face huge problems when working (finding sentiments) with the text from those languages (Cambria et al., 2018; Rahman et al., 2018; Peng et al., 2018; Satapathy et al., 2019; Milu et al., 2020; Cambria et al. 2020; Sagnika et al., 2020).

Different classification algorithms are the key to sentiment analysis. There exist many classification algorithms, for instance, naïve Bayes (NB), support vector machine (SVM), and recurrent neural network- long short term memory (RNN–LSTM, in short LSTM), provide a high level of accuracy and performance for MLSA (Altrabsheh et al., 2014; Dashtipour et al., 2016; Lo et al., 2017). However, a

customized algorithm for concept-level multilingual sentiment analysis algorithm is rare (Vilares et al., 2018; Hassan et al., 2018; Li et al., 2020; Mitra, 2020).

Datasets play a vital role in sentiment analysis. A good dataset could help a researcher to reach a solution quickly. Some existing datasets are mostly not up to date and may not be of interest (Nemes et al., 2021). Moreover, the datasets of some domains (i.e., such as student feedback) are too rare, which otherwise have a good impact on different social decision making (Nemes et al., 2021; Arican et al., 2021).

On the other hand, day by day, the approaches, techniques, and models of research on NLP (i.e., sentiment analysis) and machine learning (ML) are entering into new dimensions from heuristics to discourse structure, from coarse to fine-grained analysis, from keywords to concepts and features (Havasi, 2013; Belinkov et al., 2019). However, a concept or context-based approach is a very new dimension in processing multilingual sentiments and is not explored in depth until now (Dashtipour et al., 2016; Lo et al., 2017; Peng et al., 2017). In sentiment analysis, feature and concept extraction reduces the number of dimensions while keeping the one that better represents the large data set (Guyon et al., 2006). The feature-based sentiment analysis method mostly uses a supervised method, whereas the concept-based sentiment analysis uses unsupervised or lexicon-based methods (Musto et al., 2014). The literature shows using both of these methods and their combinations are rare (Vicent et al., 2013; Gehrmann et al., 2018; Chen et al., 2019).

The demand for preprocessing is increasing (Kumar et al., 2019) due to the massive amount of inconsistent, noisy, and incomplete data scattered throughout social media. Preprocessing is the process of removing, integrating, reducing, and transforming the data to a consistent, complete, noise-free, and understandable format (Alasadi et al., 2017). In most cases, the performance of sentiment analysis is

proportionate to the amount of preprocessing done. However, to date, a very few works have emphasized the issue of finding the best preprocessing techniques or their combinations for sentiment analysis (Krouska et al., 2016; Soni, 2017; Peng et al., 2017; Elouardighi et al., 2017; Camacho-Collados et al., 2017; Tessore et al., 2019). To be more specific, there seems to be a gap in finding optimal preprocessing techniques or their combinations for MLSA (Krouska et al., 2016; Soni, 2017; Peng et al., 2017).

1.2 Research Background

In this section, the background analyses of some of the critical terms of this research are discussed along with a brief statistic about the use of social media to clarify why this research is essential.

The statistic provided in February 2021 about the number of active users in the most popular networks worldwide (Facebook, LinkedIn, etc.) by [statistica.com](https://www.statista.com)² shows that Facebook became a market leader with near three billion registered accounts and monthly active users and possesses the first position in social media uses. Chatting site WhatsApp had over 1300 million monthly active users. Other leading apps, known as Instagram (used for photo sharing), had more than 800 million monthly users. Moreover, with more than 794 million active users, Tumblr (a blogging service) is also a famous social site.²

The individual's involvement in the social sites is highly increasing day by day, on an average of 1.56 to 2.22 hours³. They are generating vast amounts of data; some are useful (valuable data for decision-making, i.e., the data with a positive or

²<https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>

³<https://www.globalwebindex.com/reports/social>

negative review), and some are useless (the data that does not hold any information) depends on the domain of interest (Ray et al., 2015). However, only appropriate processing could derive valuable data for the intended domain. That is where sentiment analysis takes place.

1.2.1 Sentiment Analysis

Sentiment analysis (SA) is known as an application of computational linguistics, text analytics, and natural language processing that identifies and retrieves polarity from the corpus (text) by studying the provided opinions (Mihaela, 2014). SA is a tool to social network analysis and an exciting and ongoing field of study at limited length, as some of its functions cannot detect nuances (such as sarcasm, hyperbole)⁴. Even then, many resources, algorithms, datasets, and techniques are being developed to support the SA. However, due to diversity, such as multilingual and multimodal data mode, there is still scope to explore the aforementioned tools. Besides, considering both modal and multi-domain data for research is cumbersome. Also, applying all the available techniques and resources is not logical as those may not be suitable for all the applications.

1.2.2 Multilingual Sentiment Analysis

Today, most of the sentiment analysis research is done in a single language (Lo et al., 2017), mainly in English. However, as the internet grows, the diversity of communication in the web content prevails. Moreover, analyzing sentiments using one language also creates the chance to miss useful information written in other languages

⁴https://datasciencecmu.wordpress.com/category/04_social-network-analysis/

(Poria et al., 2016). Therefore, MLSA tools and techniques are developed (Boiy et al., 2009). The process of analyzing a multilingual corpus with these tools to find useful information is known as multilingual sentiment analysis. For example, after a class lecture, each student has immediate feelings about that lecture. The widely used social networks such as Facebook or Twitter become one form of shelter for the students, as they could express their thoughts to virtual friends. This expression facilitates the tutor to assess their lecture quality and improve for the upcoming classes.

However, the fact is, there are many students in a particular course and may express different thoughts on social media; also, the views are primarily expressed in their mother languages, which the researcher calls multilingual data. Therefore, if the tutor wants factual information, they should analyze those views accordingly with appropriate computerized tools.

1.2.3 Resources, Algorithms and Techniques for Sentiment Analysis

The computer cannot automatically process the different lingual data unless it has predefined lexical resources and knowledge bases (especially with concepts). More specifically, the analysis should be done using web ontologies (Lu et al., 2011). Moreover, state-of-the-art sentiment analysis methods are divided into supervised methods (Hu et al., 2013; Pak et al., 2010), which learn a classifying model based on a group of labeled data. On the other hand, unsupervised or lexicon-based methods (Ding et al., 2008; Taboada et al., 2011) infer the sentiment expressed by a part of the text based on the polarity or the intensity of the phrases or the words. The supervised methods could solve the problem associated with unsupervised methods, even though unsupervised methods are advantageous due to the massive task of labeling and

training data in a supervised method (Musto et al., 2014; Nakov et al., 2013; Rosenthal et al., 2014). However, these methods depend on some external lexical resources (i.e., SenticNet, SentiWordNet, MPQA, SentiStrength, WordNet-A) having numerical (ranges from -1 to +1) or categorical (such as positive, negative, and neutral) sentiment scores. The performances of these methods are highly reliant on the integrity and completeness of the lexical resources.

However, all these resources are mainly in English (Rahman et al., 2018; Poria et al., 2018; Satapathy et al., 2019). One such resource is SenticNet. SenticNet⁵ team has converted its resource to 40 different languages with no version for Bengali. SentiStrength (Thelwall et al., 2012) was translated automatically to address various languages, for example, Spanish. Hogenboom et al. (2014) found the sentiment scores in the Dutch version of the English version of SentiWordNet through the interrelationship between English WordNet (Miller, 1995) and its Dutch correspondent (Vossen, 1998). Ghorbel and Jacot (2011) also translated English SentiWordNet items into French. Xia et al. (2014) proposed a method for creating a version of SenticNet in the Chinese language using web dictionaries. They have adopted the English equivalent concepts in Chinese and its corresponding set of semantics. However, the resources for MLSA are still too rare, especially for the Bengali language, as there is no predefined knowledge base like SenticNet 5 (Cambria et al., 2018).

There are many classification algorithms used in the literature so far, such as linear classifier (logistic regression, Naive Bayes (NB) classifier, complementary naïve Bayes (CNB), fisher's linear discriminant), maximum entropy (ME), nearest neighbor, and SVM. Besides, decision trees, boosted trees, random forest, neural

⁵<https://sentic.net/>

networks (such as feed-forward (FF-NN), radial basis function (RBF-NN), Kohonen self-organizing, recurrent (RNN) – long short term memory (LSTM), convolutional (CNN), modular (MNN)), learning vector quantization, and quadratic classifiers. These algorithms were sometimes used individually and sometimes in combinations, sometimes for classifying based on features and sometimes based on concepts. However, the results were shown too flawed and required improvements, especially for Bengali sentiment analysis (Shirahatti et al., 2019; Banik et al., 2019; Al-Amin et al., 2017, Lo et al., 2017). Besides, it could be claimed that a general-purpose concept-level multilingual sentiment analysis algorithm is necessary to deal with the implicit meaning of data (Vilares et al., 2018; Hassan et al., 2018; Li et al., 2020; Mitra, 2020).

Datasets are the key to sentiment analysis. A good number of datasets are available such as IMDB, ABSA etc., even then, some new datasets could open the door to more research of interest (Nemes et al., 2021). One such dataset is the student feedback dataset, which will help find the students satisfaction and dissatisfaction easily and become a baseline for future research (Nemes et al., 2021; Arıcan et al., 2021).

On the other hand, more accurate sentiment could be expected once a good amount of features and concepts could be extracted (Havasi, 2013; Belinkov et al., 2019). These two techniques are thought to be a more powerful contributor of sentiments in sentiment analysis research. Languages also play an essential role. If these two techniques could be applied to any lingual data with proper combinations, the performance may have enormous improvements. (Dashtipour et al., 2016; Lo et al., 2017; Peng et al., 2017; Chen et al., 2019).

To date, many preprocessing techniques have been adopted in sentiment analysis research. Some of them are well suited to some data fields, and some are not

(Kumar et al., 2019). So, identifying appropriate preprocessing techniques is field dependant. Moreover, the performance of sentiment analysis is directly connected to the amount of preprocessing done. A recent survey shows improved performance for sentiment analysis with preprocessing technique combinations (Krouska et al., 2016; Soni, 2017; Peng et al., 2017; Elouardighi et al., 2017; Camacho-Collados et al., 2017; Tessore et al., 2019).

1.3 Problem Statement

From the background and literature review, it is evident that today's market is very competitive. Analyzing and predicting public opinion (especially student feedback) is very important to survive in the market. There are many possibilities for students to turn over if the students' problems are not adequately noticed. Meanwhile, the students' expressions on many social media pages (e.g., Facebook page, etc.) regarding their institutions are in different languages such as English, Bangla, Spanish, Arabic, Chinese, etc. Therefore, noticing those expressions accordingly and taking proper action is crucial, especially processing them automatically in real-time is of more importance.

Researchers have provided recent surveys that multilingual sentiment analysis has not been entirely investigated and hold enormous possibilities as an application (Cambria et al., 2018). In particular, research showed that future work includes an investigation of different knowledge bases, polarity lexicon, algorithms, preprocessing, and feature or concept extraction techniques and their effect on model performance (Altrabsheh, 2014; Sindhu et al., 2019; Jamatia et al., 2019). Moreover, most (Feldman, 2013; Poria et al., 2018; Saxena et al., 2019) of the analysis tools do a

single-term investigation on the document only. However, the concept-level investigation task with the use of knowledge base achieves accurate testing of concepts on the sentence, document, and corpus levels rather than a single-term investigation on the document only (Manwatkar et al., 2011; Feldman, 2013; Cambria et al., 2016; Poria et al., 2018; Cambria et al., 2020). Background and literature also show that there is a lack of lexical resources (i.e., knowledge base and polarity lexicon) for multilingual (especially Bengali) sentiment analysis at the concept level (Rahman et al., 2018, Cambria et al., 2018, Milu et al., 2020; Cambria et al., 2020). As a result, there requires Bengali knowledge base and polarity lexicon to promote Bengali sentiment analysis research.

Literature review reveals that there are many machine algorithms used so far for this sort of analysis (Altrabsheh et al., 2014; Dashtipour et al., 2016; Lo et al., 2017; Al-Amin et al., 2017; Banik et al., 2018; Shirahatti et al., 2019). These algorithms are found suitable for feature-level sentiment analysis. Besides, some concept-level sentiment analysis also shown good performance, especially for English lingual data. However, due to knowledge bases' insufficiency, these algorithms were not tested at the concept-level with Bengali data. Also, a recent survey shows a gap of research on concept-level multilingual sentiment analysis algorithms and has enormous scope to explore it by proposing a new algorithm (Vilares et al., 2018; Hassan et al., 2018; Li et al., 2020; Mitra, 2020).

One of the essential resources in sentiment analysis is appropriate Datasets. The literature and background study shows, a well-formed dataset could help a researcher with improved model performance. A good number of datasets are available such as IMDB, ABSA etc. However, these datasets are mostly not up to date and may not be of interest (Nemes et al., 2021). Research shows, there is a lack of

student feedback datasets on Bengali and English lingual data and poses good scope of creating new datasets for the said domain to help in finding the students satisfaction and dissatisfaction quickly and become a baseline for future research (Nemes et al., 2021; Arican et al., 2021).

The state-of-the-art literature shows almost every sentiment classification task requires feature extraction. Existing studies (Shah et al., 2016; Ahmad et al., 2019; Ganesh et al., 2019; Cambria et al., 2020) have emphasized either on feature-based SA or concept-based SA. The studies ignored using these two extraction techniques in combination and applying them in the language like Bengali. The studies indicate that the feature and concept extraction technique is not a single criterion for performance enhancement, but also adopting a proper combination of feature and concept extraction techniques are the critical criteria for performance enhancement ((Nguyen et al., 2018; Cabada et al., 2018; Shirahatti et al., 2019, Bhargava et al., 2019; Jamatia et al., 2019; Sazzed et al., 2019; Cambria et al., 2020). Therefore, analyzing performance with different features and concept extraction techniques and combinations is necessary to enhance current research's performance and guide future research.

The state-of-the-art literature shows almost every sentiment classification task requires preprocessing. The study shows that some works exist where the contributors tried to find the best preprocessing techniques. However, the literature indicates that those studies (Krouska et al., 2016; Soni, 2017; Peng et al., 2017; Elouardighi et al., 2017; Camacho-Collados et al., 2017; Tessore et al., 2019; Milu et al., 2020) ignored preprocessing technique combinations and mainly applied to monolingual data. The overall performance of different investigations shown in the literature review indicates, preprocessing is not only the criteria for performance enhancement but also

adopting a proper combination of preprocessing techniques is the critical criteria for performance enhancement (Raiyani et al., 2018; Sindhu et al., 2019; Taher et al., 2018; Bhargava et al., 2019; Jamatia et al., 2019; Zobeidi et al., 2019; Sarkar, 2019; Milu et al., 2020). Thus, examining different preprocessing techniques with combination to evaluate the optimal one is necessary to improve current research performance and for future reference in researching solutions efficiently.

1.4 Research Questions

The following research questions were investigated in this study to meet the problems stated in the problem statement section:

1. What are the methods of creating Bengali knowledge bases or polarity lexicons?
2. Which algorithms of multilingual concept level sentiment analysis are useful? Proposed or baseline algorithms (NB, SVM, and LSTM)?
3. How to address the lack of Bengali student feedback datasets?
4. Which extraction techniques (feature and concept) and their combinations are the best in the scale of performance in multilingual sentiment analysis?
5. Which preprocessing techniques and their combinations are optimum in processing multilingual sentiments?

1.5 Research Objectives

This research's prime objective (RO) is to search for optimal techniques and resources for Bengali and English lingual data. However, as per the literature (in chapter 2), the problems statement, and the research questions, this prime objective

could be met by ensuring the following objectives. Therefore, the objectives of this research are:

- 1. To create a Bengali knowledge base and polarity lexicon about the English knowledge base and polarity lexicon SenticNet 5**
- 2. To propose an algorithm for concept-level multilingual sentiment analysis.**
- 3. To create well-form Bengali and English datasets on students feedback**
- 4. To analyze and evaluate the best among different feature(s) and concept extraction techniques with their combinations for multilingual sentiment analysis.**
- 5. To examine the preprocessing techniques with different combinations for multilingual sentiment analysis and search optimal one.**

1.6 Research Significance

This study attempts to help those researchers who are working on natural language processing (NLP), especially on multilingual data such as English and Bengali. This study will also help the researchers to quickly understand the tools (preprocessing, feature or concept extractions, machine learning algorithms, datasets, polarity lexicon, and knowledge bases) they need to adopt for this sort of analysis. The study will help the Bengali researchers quickly proceed with sentiment analysis using BanglaSenticNet- a knowledge base developed in this research.

This study will assist future research in the concept-level sentiment analysis (CLSA) with the adaptation of proposed architecture and algorithms. This research will be helpful to those institutions or organizations that are dependent on the web or social media feedback of their incumbents in advancing their business or ideas. As the

study has considered data of student feedback in this thesis, the study will therefore, help both the students and teachers in advancing the research in the area of education. More specifically, this research will benefit the entire industry and the nation in advancing several different areas similar to this study through further research.

1.7 Research Scope

Firstly, this research has considered Facebook data of students' feedback as the Facebook messages are not limited in size like tweets (Troussas et al., 2013), and this domain has the possibility of further research (Niu, 2019). Secondly, this research has considered English and Bengali data as the aim is to add novel contributions (i.e., knowledge bases, lexicons, and datasets) to the research on these languages. The time duration of data collection was limited to one year, as sufficient data were available in this period. The beneficiaries to this research in further studies are students, instructors, educational institutions, and NLP researchers.

From the list of polarity lexicon and knowledge bases such as SenticNet, AffectNet, and ConceptNet, etc. this research used SenticNet version 5.0 (as it contains a very rich set of affective and conceptual information related to SA (Bisio et al., 2017; Ho et al., 2018; Cambria et al., 2018)). An additional knowledge base (BanglaSenticNet) and polarity lexicon were developed for the Bengali language and used in this research.

The classification algorithms used in this study are NB, SVM, and LSTM, as the researchers proved these algorithms provide a high level of accuracy and performance (Jagdale et al., 2019; Pouyanfar et al., 2019; Gamal et al., 2019; Satapathy et al., 2019). An additional MLSA algorithm (MCSAlgo) was proposed and

applied in this research. This study has used two extraction approaches in the analysis such as a feature-based (i.e., TD - IDF (Simple), unigram, bigram, trigram, parts of speech) approach and concepts-based (i.e., concepts) approach as they are proven as better on the scale of classification accuracy (Abualigah et al., 2017; Bisio et al., 2017; Jing, 2017; Satapathy et al., 2019; Gamal et al., 2019; Zobeidi et al., 2019).

The preprocessing techniques adopted in this research are, remove punctuation (RP), negation (N), reduction of letter reputation (RLR), stop word deduction (SWD), stemming (S), tokenization (T), and case conversion (CC). The literature (Ghosal et al., 2015; Krouska et al., 2016; Elouardighi et al., 2017; Rahman et al., 2019; Sarkar, 2019) shows, applying these preprocessing techniques improves the performance (i.e., classification accuracy) of the sentiment analysis.

1.8 Research Contribution

Every research is conducted to ensure some new knowledge being added to the field of study. The contribution of this thesis is listed below:

Firstly, this thesis has contributed a Bengali knowledgebase (BanglaSenticNet) with 30000 concepts and near about 150000 semantics of those concepts. Besides, this research has contributed a Bengali polarity lexicon with 72433 concepts to help the researcher in this field in the expansion of research in the future. This Bengali knowledge base and polarity lexicon are novel and first of its kind to the best of the author's knowledge.

Secondly, this research has contributed an algorithm (MCSAlgo) for MLSA at the concept-level; this will help the researchers analyze the text from different languages using concepts. Also, recommended more accurate performance

measurement algorithms for MLSA from among NB, SVM, LSTM, and MCSA (proposed algorithms).

Thirdly, this research has contributed three well-form datasets (two in English and one in Bengali) that could be further investigated and help related organizations when needed. Here, well-formed means, collected noisy raw data are well processed with the preprocessing techniques mentioned above, such as RP, N, RLR, SWD, S, T, and CC.

Fourthly, this research has presented the best extraction techniques among TD - IDF (Simple), unigram, bigram, trigram, parts of speech, and concepts for MLSA using machine learning algorithms such as NB, SVM, and LSTM to help the researchers conclude quickly.

Fifthly, this research has presented optimal preprocessing techniques (RP, N, RLR, SWD, S, T, and CC) with their combinations for MLSA intending to help the researchers save their preprocessing time.

1.9 Summary

The purpose of this research is to find the optimal techniques (preprocessing, feature or concept) and resources (datasets, knowledge bases, and lexicons) for sentiment analysis from Bengali and English lingual data (i.e., student's feedback) using machine learning, deep learning and proposed algorithm. Therefore, this chapter discussed some of the critical terminology of this research to make clear why this research is essential.

Analyzing the increasing amount of multilingual information (i.e.; student feedback) in the social media regularly using sentiment analysis (SA) techniques and

resources could identify emotions of the student's towards the current teaching, i.e., either positive or negative. However, commonly used resources and techniques for English are very rarely explored and tested for some other languages such as Bengali. Also, there is a lack of own resources and techniques of SA for those languages. In such a context, the background section of this chapter gave the glimpse of resources, techniques as well as algorithms used in the literature for different languages especially, English and Bengali along with basics of SA and MLSA.

The background section elucidates different gaps for MSA such as 1) lack of optimal preprocessing techniques or their combinations, 2) lack of optimal feature and concept extraction techniques and their combinations, 3) Inappropriate use of machine learning, deep learning algorithms, 4) rare collection of resources such as datasets, knowledge bases, and lexicons. To meet this gap, this research has posed five questions and is presented in the research questions section of this chapter.

The research is guided by research questions and objectives. To meet different gaps poses by background and different research questions of the study; this research is guided by five research objectives. The objectives are creating knowledge resources and evaluating them with state-of-art resources, proposing an algorithm for concept-level multilingual sentiment analysis, creating new datasets in both Bengali and English, examining the preprocessing techniques and their combinations for optimal one, analyzing and evaluating the best extraction techniques and their combinations. These objectives are discussed in the research objective section of this chapter.

Defining research scope is an important part of any research and as per general practice; the section is included in this chapter. This research used multilingual (English and Bengali) student's feedback of one year from Facebook. The study has adopted the lexicon and knowledge bases such as SenticNet 5 and BanglaSenticNet.

The research then used different algorithms such as NB, SVM, LSTM and MCSAlgo (proposed algorithm) to evaluate the techniques and resources.

In addition, the research has extracted different features/concepts such as TD-IDF (Simple), unigram, bigram, trigram, parts of speech, and concepts. The research also conducted different types of preprocessing such as remove punctuation, negation, reduction of letter reputation, stop word deduction, stemming, tokenization and case conversion. The research significance is also included in this chapter. This study is significant to NLP researcher especially Bengali, to quickly understand the tools and proceed with multilingual concept-level sentiment analysis. This research has mainly five contributions in accordance with objectives; such as a Bengali knowledgebase (BanglaSenticNet) with 30000 concepts with near about 150000 semantics of those concepts and Bengali polarity lexicon with 72433 concepts; an algorithm (MCSAlgo) for MLSA at concept-level. In addition, three well form datasets (two in English, and one Bengali), that could be further investigated and helps the related organizations if needed. Besides, optimal preprocessing techniques and the best extraction techniques were evaluated for multilingual sentiment analysis.

Each of the contribution is described in details and included in this chapter. The next chapter provides details and a comprehensive review of related literature to give a bigger picture of the field of study and justification of this study.