A MULTI-ASPECT BASED OPINION MINING SYSTEM FOR OPEN AND DISTANCE EDUCATION USING ONLINE REVIEWS

Ashok Kumar J
Research Scholar
Department of Information Science and Technology
CEG Campus, Anna University
Chennai – 600025
Email: jashokkumar83@gmail.com, jashokkumar83@auist.net

Abirami Murugappan
Assistant Professor
Department of Information Science and Technology
CEG Campus, Anna University
Chennai – 600025
Email: abirami.murugappan@gmail.com, abirami@auist.net

ABSTRACT
Open and Distance education is the form of education that delivers pedagogy, technology, and instructional designs to students who are not physically available at the same place in a traditional classroom or campus. Opinion mining (also called sentiment analysis) plays an important role in the field of social media. It computes people's polarities such as positive, negative, neutral, which were expressed in online social media contents at various levels, namely, document level, sentence level, and corpus level. In this research paper, a multi-aspect based opinion mining system is proposed by applying opinion mining techniques for Open and Distance Education social media contents. The purpose of this research is to measure the public satisfaction of open and distance at the title level, document level, sentence level, and aspect level. The proposed system was employed by the data collection process, preprocessing, feature extraction, opinion detection and polarity classification using the Naïve Bayles classifier. The detected opinions at various levels are also visualized. The performance of the system is evaluated using precision, recall, f-measure, and accuracy.
Keywords: Distance education, opinion mining, sentiment analysis, visualization, open education

INTRODUCTION
The first distance education course was started by Sir Isaac Pitman in the year 1840s in the way of sending and receiving shorthand texts to students on postcards. This success laid the foundation to start Phonographic correspondence society to offer formal courses. Then the first correspondence or distance learning degree was offered by the University of London. Later, the Open University was founded to create a respectable learning from the traditional form of education. According to Vasileios Kagklis et al. (2015), the distance education is a form of education in which there is a limited interaction between teachers and students. M.Banu Gündoğan and Gülsün Eby (2012) defined the distance education as a result of the struggle for permanence and progress. Safiullin Lenara et al. (2014) defined the distance learning is the independent form of education by using pedagogical technologies to design and implement remote courses. Distance education system is needed by various communities, namely, pupils in rural areas, small cities, managers, army officers, etc. Gao Guohong, Li Ning, Xiao Wenxian, and Wan Wenlong (2012) stated that the modern distance education is established using computer networks and its features, namely, multimedia teaching, interactive demonstration, classroom management, online examinations and so on. This form of education provides the learners to get more information using the internet. The modern distance education is implemented by both synchronous and asynchronous methods. Synchronous method is a face to face communication for sharing data and information between two or more computers. Examples, written texts, audio tools, and video, etc, (Tahir Tavukcu, øbrahim Arap, & Deniz Özcan, 2011). The asynchronous method delivers the course contents in one way like books, CD-ROMs, and videotapes. Most of the working professionals, housewives, Government officials, etc., like to pursue distance education to upgrade their knowledge or to get expertise in that particular course or to get promoted in the employment service. There are many Universities, centers, and private institutions offer distance education courses in each and every country. People might be asking opinions from friends, family members, and educators about the University, course details, course fee, and validity of the degree before getting admission. Nowadays, people are expressing their views, experiences, and opinions about the Universities through online social media contents. Opinion mining plays an important role in the field of social media to get people's positive or negative or neutral sentiment. E.g., great, amazing, very nice, wonderful, bad, and poor. These opinions can be expressed in online feedback forms, web emails and social networking websites such as Facebook, Twitter, LinkedIn, YouTube, MySpace, Blogs, and forums, etc., about the product sales, products service, quality, policy initiatives, Institutions, forecasting political opinions and news contents. Opinion mining

(also called sentiment analysis) computes these sentiments at the document level, sentence level, and corpus level. This computation helps the consumer and the public to ensure their opinions in social media.

In this paper, we analyzed open and distance education of Indira Gandhi National Open University (IGNOU) in India by using people's online reviews. IGNOU is the Public, Central University established in the year 1985. It is the largest university in the world with over 4 million students to impart open and distance education. The University offers higher education opportunities in India, particularly to the disadvantaged segments of society to strengthen the human resources. The Commonwealth of Learning (COL) designated that IGNOU is the first centre of Excellence in open and distance education in the world. In this context, we propose a multi-aspect based opinion mining system for Open and Distance Education to measure the satisfaction of IGNOU. In this research paper, first, the people's online reviews are obtained about IGNOU from social media websites. Second, the data pre-processing method is applied to clean the data. Third, features are extracted from the cleaned data. Fourth, opinions are detected for the extracted features. Fifth, the detected opinions are visualized. Finally, the Naïve Bayes classifier is employed to predict the accuracy of the actual and proposed system. The rest of the paper is organized as follows. Section 2, explained the related works in open and distance education and opinion mining. Section 3, explained the proposed opinion mining system for open and distance education. Section 4, described the experimental results and discussion. Finally, the conclusion and future works are presented in section 5.

RELATED WORKS

Open and Distance Education

The open and distance education is a system of learning and teaching. It disseminates the knowledge and quality higher education through open and distance learning mode. Ebru Melek Koc (13) studied the roles of mentors such as self-trainer, networker, social supporter, academic support and psychological supporter in a distance learning teacher education program. Yingqi Tang (2013) analyzed the librarian position announcement in the context of distance education librarianship in the United States. The authors identified the fundamental occupational skills, namely, technology skills, information science skills, and communication skills. Vasileios Vasileios Kagklis et al. (2015) studied the online discussion of postgraduate students at the Hellenic Open University by using text mining techniques and sentiment analysis. The authors provided valuable information to improve the educational process. Safiullin Lenara, Fatkhiev Arturb, Saipullaev Ullubic, and Bagautdinova Nailya (2014) proved the distance education is a new form of education by using the pedagogical technologies. Perihan Paksoy [16] studied the developments of Turkish airlines by using a global distance education program. Hieronymus J.M. Gijselaers et al. (2015) investigated the sedentary behavior and physical activity to predict study progress in distance education. The authors suggested that sedentary behavior contributes a significant predictive for learning performance in distance education. Fatma Kübra Çelena, Aygül Çelikb, and Süleyman Sadi Seferoğlu (2013) analyzed teachers approaches to distance education. The authors showed that the majority of participants want to participate in distance learning activities. They also reported that half of the participants don't trust distance education systems operated in Turkey in terms of content, materials used, evaluation methods, the validity of certificates and career opportunities. Filiz Kanteka (2014) presented the current status of distance education in nursing. This program attracts many nurses who are working in public and private institutions. Annelien van Rooyen (2015) studied the perception of accounting students in distance education through social media at the University of South Africa. The authors suggested that accounting modules need to increase success rates by using these social media apps to provide lectures. Bahar BERBEROĞLU (2015) evaluated the policy of open and distance education programs of Anadolu University by using linear models. This work was implemented by considering the relationship between time and number of programs. The authors reported that the numbers are increased for open and distance education in 1993 and in 2009 due to jumps in the graphics. Nurmukhametov N, Temirova A, and Bekzhanova T (2015) studied the organization and management of distance education in Kazakhstan by using Internet technologies. Galina Samigulina and Zarina Samigulina (2016) proposed distance education intelligent system based on the biological approach to Artificial Immune Systems (AIS) for the establishment of a multilateral exchange of information. The authors analyzed that AIS allows individual learning to gain necessary skills in real time based on modular training courses and accessing modern equipment.

Opinion Mining

Opinion mining is the effective method to measure public opinion about the open and distance education. Oksan Bayulgen and Ekim Arbatli (2013) investigated the Cold War rhetoric in the US – Russia relations by looking at the 2008 Russia – Georgia war based on content analysis and public opinions, links between media, public opinion, and foreign policy. Pawel Sobkowicz, Michael Kaschesky, and Guillaume Bouchard (2012) introduced an opinion formation framework based on content analysis of social media and socio-physical system modeling by automated topic, emotion and opinion detection in real-time, information flow modeling and agent-based simulation and modeling of the opinion of networks. Magdalini Eirinaki, Shamita Pisal, and Japinder Singh (2012) presented an opinion search engine system with two novel opinion mining algorithms. First, the high adjective count algorithm was used for identifying and extracting features that are deemed as the most important and characteristic of each review. Second, the max opinion score algorithm was used to assign ranks to the features for deciding the final classification of reviews such as positive, neutral or negative. Daniel E. O' Leary (2011) captured opinion expressions in blogs and determined its opinions at different levels, namely, a word, sentence, and paragraph. The author used mood declaration approach to identify positive, negative or neutral opinions, and opinion word approach to identify specific words that suggest a particular opinion, and domain-specific approach to improving the quality of analysis. Andres Montoyo, Patricio Martinez-Barco, and Alexandra Balahur (2012) employed the tasks of subjectivity analysis and sentiment analysis to detect private states (opinions, emotions, sentiments, beliefs, speculations) on different topics. The authors grouped four categories, each with their corresponding challenges such as the creation of resources, classification text, opinion extraction and applications of sentiment analysis. Farhan Hassan, Saba Bashir, and Usman Qamar (2014) proposed a new algorithm for twitter opinion mining (TOM) to improve the accuracy of text classification and resolve the data sparsity issues. This algorithm was employed based on data acquisition, preprocessing, and polarity classification algorithm and evaluation procedure. Deanne K. Bird et al. (2014) conducted an online survey for Nuclear Power in Australia in 2010 and 2012. The study examined, comparatively analyzed and assessed public opinions (positive or negative or neither) by a set of questionnaires regarding climate change and the Fukushima disaster.

A. Moreo, M. Romero, J.L. Castro, and J.M. Zurita (2012) developed a lexicon-based system to analyze user comments on current news items in social media. The system identifies the discussion topics on which user expressed their opinions. The authors studied the system analytically with the preparation of knowledge (filtering stage, preprocessing stage), opinion focus detection module (interpretation context, disambiguation analysis, and frequency analysis), sentiment analysis module (labeling expression stage, tuples extraction stage, tuples clustering and filtering) and sentiment mining module. IT Vendors routinely use YouTube as one of the tools in social media to disseminate their IT product information to acquire customer opinions. In this scenario, the authors applied deep sentiment analysis for identifying and extracting value structure polarity (writer's opinion), attitude polarity (number of opinions about the source materials), the level of intensity (degree adverbs), and automated personality-sentiment-value analysis procedure for marketing strategies (Haeng-Jin Jang, Jaemoon Sim, Yonnim Lee, & Ohbyung Kwon, 2013). Xiao Wang (2013) examined the motivations and factors to predict one's intention to use social media while viewing mediated sports events by applying the integrative model of behavioral prediction and attitude functions. The author conducted a web-based survey in February 2012 among the limited number of students with a set of the questionnaire when they watch mediated sports events. In this model, structural equation modeling analysis and controlling were employed to predict intentions to use social media. Qiwei He et al. (2014) evaluated the relationship between posts and self-monitoring (SM) skills based on Snyder's SM Questionnaire (1974) collected via the internet. The authors introduced an Item Response Theory (IRT) model to check the validity of the Internet data for online assessments and predicted users SM skills by using both structured and unstructured textual analysis.

THE PROPOSED SYSTEM

An opinion mining system is the most effective way to measure the public opinions that are expressed in any form or language. The multi-aspect based opinion mining system applies various steps for analyzing online reviews. The system involved in the data collection process, data preprocessing, feature extraction, opinion detection at the title level, document level, sentence level, and aspect level, opinion visualization, opinion classification using a Naïve Bayes algorithm, and performance evaluation using precision, recall, f-measure, and f-measure. The general architecture of the proposed system is shown in Fig.1.

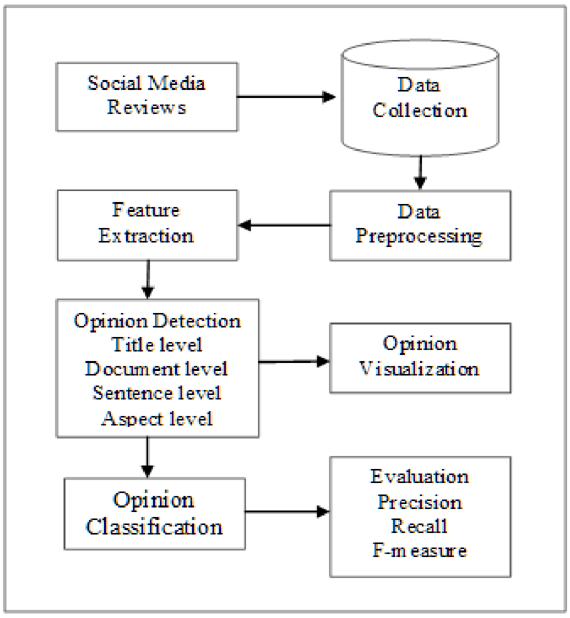

Figure 1. A multi-aspect based opinion mining system

Data collection

Data collection is the process of collecting information on the specified topic or targeted variables. The data can be a quantitative data or qualitative data. It is used to evaluate outcomes to answer relevant questions in terms of a research or decision-making process. There are many methods used to collect freely available information in online social media. Initially, the data are collected in the ways of surveys, personal interviews and focus groups. Later, the data are collected using various web analytics tools. These tools collect data in the form of structured, semi-structured and unstructured formats. The largest social media websites (Facebook, Twitter, Blogs, etc,.) are using APIs to collect accurate data. In this paper, the reviews are collected from mouthshut website about IGNOU using python programming. The data considered only from January 2016 to September 2016. It contains 102 online reviews from different people.

Data preprocessing

Data preprocessing is an important step in the field of opinion mining and sentiment analysis. The real world data are incomplete, noisy, and inconsistent. It doesn't make any sense to analyze those data. The quality of data is the most important factor in analyzing data. If there is much incomplete, noisy and inconsistent data, Then the feature extraction or knowledge discovery is more difficult. In this step, URLs, profile images, profile links, send message labels, title values, data time contents, data time links, review detail links, font values, and font links are removed to obtain profile names, genders, review titles, and review summaries.

Feature extraction

Feature extraction is the process of transforming arbitrary data into numerical or string features for machine learning. Features are also inherent properties of the data. These features are an initial set of measured data, which are used for dimensionality reduction. If there is too much of information, then it reduces the effectiveness of the knowledge discovery or data mining. In this approach, the collected review contains more information, namely URLs, profile images, profile links, profile names, send message labels, gender details, title values, data time contents, data time links, review titles, review detail links, font values, font links, and review summaries. If this information is used for opinion mining and sentiment analysis, then the performance of the system will be not effective. In this context, only the titles, text reviews, sentences, and aspects are considered as the feature set for the opinion mining system. The cleaned data are processed and extracted theses feature sets using python programming. For instance, the review, "<title>Best faculty</title><p>*Indira Gandhi national open university(IGNOU) is a best one university in India for the dropper, for privateness. And one quality more of this university is that this gave the opportunity to every student to complete his/her study with no burden. Because many universities are said to his/her students to get rid of the university in case you couldn't qualify your class in 2-3 years but in this case, IGNOU is the best university and much more other great qualities makes it great*.</p>" is segmented into titles and review texts. Further, the review texts are segmented into sentences. These sentences are processed to extract aspects.

Opinion detection

Opinion mining is the process of computing people's opinion on posts, reviews, comments, news, etc., which are expressed in Blogs, Forums, Facebook, Twitter, YouTube, etc, In this model, MeaningCloud is used to detect opinions for the extracted feature set. The MeaningCloud performs various operations, namely, text classification, sentiment analysis, Language identification, topic extraction, and text clustering using an Application Programming Interface (API) Key. The sentiment analysis operation measures the polarity of each feature set as positive, negative and neutral. The proposed approach detects sentiment at the title level, document level, sentence level, and aspect level. Mathematically, $(T, R) = \{(t_1, r_1), (t_2, r_2), (t_3, r_3), \ldots, (t_n, r_n)\}$ where T and R represents the association between title and review respectively. $t_1, t_2, t_3 \ldots t_n$ represents titles associated with reviews. $r_1, r_2, r_3 \ldots r_n$ represents reviews associated with titles. Further, $(R, S) = \{(r_1, (s_1, s_2, s_3 \ldots s_n)), (r_2, (s_1, s_2, s_3 \ldots s_n)), \ldots, (r_n, (s_1, s_2, s_3 \ldots s_n))\}$. Where, R and S represent reviews and sentences of the reviews respectively. r_1 represents first review and $s_1, s_2, s_3 \ldots s_n$ represents sentences of the first review, and so on.

Title level opinion detection

A title is the first impression of a review. Normally, people look at the title first instead of going through the entire review. Reviews are written based on titles. The titles are very important to understand the interest and disinterest of people. Therefore, opinions are detected as positive, negative, and neutral at the title level.

Document level opinion detection

The opinions expressing at the document-level is not a single opinion. It is based on multiple opinions. The document level opinion mining aims to classify reviews as positive, negative, and neutral. For instance, "*Indira Gandhi National Open University based in Delhi is a very good option for those who are doing their job or preparing for something and at the same time want to complete their education. The best thing is that you can go for the exam in any month with your convenience. The syllabus is normal and it provides books or you can also get it online. The fee is normal in our range. The reputation of this university is quite good. You can learn from anywhere, at any age with your convenience. All types of 10+ 2 courses, degree courses, etc. are available. If you are planning to study with your job, then it is the best option available, just go for it.*" All sentences in this review expressed as positive except the sentence "*The fee is normal in our range*". Therefore, the above-mentioned review expresses an overall opinion as positive. Similarly, opinions are detected for all reviews at the document level.

Sentence level opinion detection

The sentence level opinion detection determines each sentence as positive, negative, and neutral. This level of opinion detection is closely related to subjectivity sentiment classification, which expresses subjective views and opinions of the opinion holder (Bing Liu, 2012). The opinions are detected for simple and compound sentences. The overall score is calculated by adding the maximum polarity and minimum polarity. For instance, the sentence "*during the admission, it was in my mind that it will be good to do a distance course or it will be a setback for my future but on being a part of IGNOU*" express both positive and negative opinions. For "admission" the sentence is positive, but for "future", the sentence is negative. The overall polarity of the sentence is detected as neutral. Similarly, sentiment polarity for each sentence is detected.

Aspect level opinion detection

The opinions expressed at the title level, document level, and sentence level do not express individuals likes and dislikes on opinion targets, The opinion targets are entities and its aspects (or attributes). This target expresses individuals and groups opinions exactly. At this level of analysis, the named entities and abstract concepts are identified. Then, a polarity value (positive, negative, neutral) is assigned to those entities and aspects. For instance, the sentence "*IGNOU is a very good university for those who don't want to go college or for those who are unable to go college*" is processed and extracted the entity "IGNOU" and aspect "University". The polarity of both entity and aspect are identified as positive opinion.

Visualization of opinions

Visualization is one of the best methods to communicate a message or knowledge information by creating images, diagrams, and animations. This makes an effective way to explore abstract ideas. Visualization techniques are used in many applications, namely, science, engineering, education, medicine, multimedia, etc. In this research, the extracted opinions at the title level, sentence level, document level and aspect level are visualized using Nodexl as shown in Fig. 2 - 5. In the visualization, first, the opinions, namely, P+, P, NEU, N, N+, and NONE are highlighted by blue color with a primary label (degree of the vertex) and secondary label (name of the vertex) in title level opinion detection. Second, the green, maroon, purple, fuchsia and blue colors are highlighted in document level opinion detection except for the opinion N+. Third, the yellow, green, red, blue, olive, and violet colors are highlighted in sentence level opinion detection as well as aspect level opinion detection.

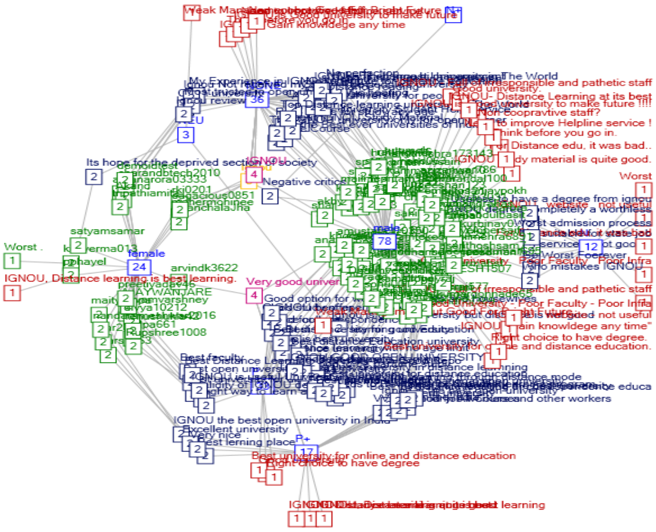

Figure 2. Visualization of title level opinion detection

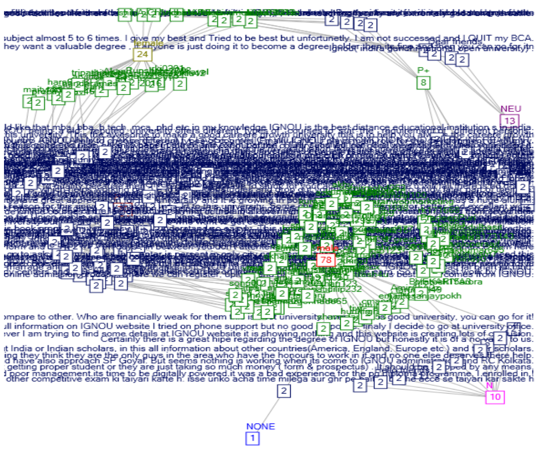

Figure 3. Visualization of document level opinion detection

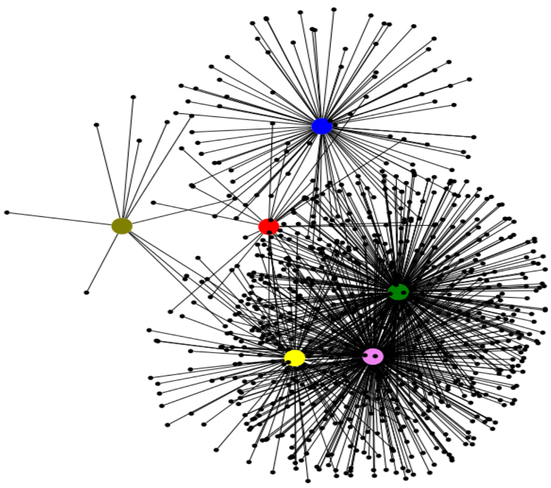

Figure 4. Visualization of sentence level opinion detection

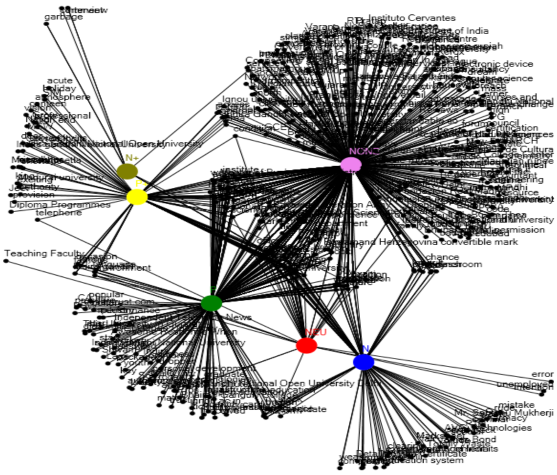

Figure 5. Visualization of aspect level opinion detection

Naïve Bayes Classifier:
In machine learning, Naive Bayes classifier is a probabilistic model based on Bayes' rule (Eq. 1) and a set of conditional independence assumptions. This classifier is used to assign the most likely class for each level, and it performs well for sentiment classification problems. The Naïve Bayes algorithm is fast and space efficient. It is also described as an independent feature model.

$$P(c_j|d) = \frac{P(c_j)P(d/c_j)}{P(d)} \tag{1}$$

Where $P(c_j|d)$ refers the probability of an event d in class c_j, $P(d|c_j)$ refers the probability of generating events d d in class c_j, $P(c_j)$ refers the probability of occurrence of class c_j, and $P(d)$ refers the probability of event d occurring.

The documents are generated as a multinomial event model by considering word frequency information. A document is represented as a bag-of-words model. The classifier checks the presence of a positive or negative or neutral opinion in each feature set. If the polarity word appears in the document, then the score is updated as one The probability of an event is obtained in a document based on equation (2).

$$P(d_i|c_j) = P(|d_i|)|d_i|!\prod_{t=1}^{|V|}\frac{P(w_t|c_j)^{N_{it}}}{N_{it}!} \tag{2}$$

Where $P(d_i|c_j)$ is represented as the probability of i^{th} document belong to the j^{th} class. $P(|d_i|)$ represents the probability of i^{th} document. $P(w_i|c_j)$ represents the probability of occurrence of i^{th} word in j^{th} class. N_{it} is the total occurrence of a polarity word t in the document d_i. The classification process is optimal for selecting the highest probable class with dependent features.

EXPERIMENTAL EVALUATION AND DISCUSSION
The open and distance education dataset was collected from online social media using the data collection process. This dataset is used to carry out the experiments at four levels, namely, title level, document level, sentence level, and aspect level. The features were extracted from the dataset as 102 titles, 102 documents, 783 sentences and 1683 aspects. Then, the opinions were detected by processing all these features using a MeaningCloud tool (http://www.meaningcloud.com). At each level, the opinions were detected as highly positive (P+), positive (P), neutral (NEU), negative (N), highly negative (N+), and objectives (NONE) as shown in Table 1. The objective statements have not expressed any opinions or meaningful statements. For experimental purposes, the highly positive and positive opinions are considered as a single positive class. Similarly, the highly negative and negative opinions are considered as a single negative class (Table 2). The neutral class has been used as it is detected. The objective statements are not considered for the experiment. Naïve Bayes classifier has been used to perform the classification task using the Weka data mining tool (Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten, 2009) at 10 cross fold cross-validation. First, the classification was performed at each level as shown in Table 4. Second, the classification was performed by combining all the feature sets, and performance of the system is predicted by

detailed accuracy in each class as shown in Table 5. The proposed system is evaluated using confusion matrices, precision, recall, f-measure, and accuracy.

Table 1. Opinion detection

Features	P+	P	NEU	N	N+	NONE	Total
Title level	17	33	3	12	1	36	102
Document level	8	70	13	10	0	1	102
Sentence level	77	288	22	87	11	298	783
Aspect level	145	417	43	107	11	960	1683

Table 2. Combining opinions as a single class

Features	P	NEU	N	Total
Title level	50	3	13	66
Document level	78	13	10	101
Sentence level	365	22	98	485
Aspect level	562	43	118	723
Combined Method	1055	81	239	1375

A three class confusion matrix is defined as shown in Table 3. The confusion matrix is also called as contingency table or an error matrix. It contains the classification result about actual and predicted classifications. The acutal class is represented in rows and predicted class is represented in columns. In the confusion matrix, A, B, and C are represented as a class, namely positive (P), negative (N), and neutral (NEU) respectively. The diagonal entires tpA, tpB, and tpC represent the number of correctly classified data for each class as positive, negative, and neutral. The remaining entries represent incorrectly classified data for each class. Precision is the fraction of predicted positive cases that were correct. The recall is the fraction of positive cases that were correctly identified (Howard J. Hamilton, 2012). Precision, Recall, F-measure, and accuracy are defined as follows. If there is high precision, then the system returns more relevant results. If there is a high recall, then the system returns most of the relevant results. The f-measure returns the weighted average of the precision and recall.

Table 3. Confusion Matrix

		Predicted		
		A	B	C
Actual	A	tpA	eAB	eAC
	B	eBA	tpB	eBC
	C	eCA	eCB	tpC

$$\text{Precision } A = \frac{tpA}{tpA + eBA + eCA}$$

Where tpA is the number of true positive predictions for the class A and eBA, eCA are false positives.

$$\text{Recall } A = \frac{tpA}{tpA + eAB + eAC}$$

Where tpA is the number of true positive predictions for the class A, and eAB, eAC are false negatives.

$$\text{F-measure} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

$$\text{Accuracy} = \frac{2 \text{ x } (\text{TruePositive} + \text{TrueNegative} + \text{TrueNeutrals})}{\text{TruePositive} + \text{FalsePositive} + \text{TrueNegative} + \text{FalseNegative} + \text{TrueNeutrals} + \text{FalseNeutrals}}$$

Table 4. Classification result using confusion matrices

Feature level		P	N	NEU	Precision (%)	Recall (%)	F-measure (%)	Accuracy
Title level	P	50	0	0	75.75	100	86.20	75.76
	N	13	0	0	0	0	0	
	NEU	3	0	0	0	0	0	
Document level	P	78	0	0	77.22	100	87.15	77.23
	N	10	0	0	0	0	0	
	NEU	13	0	0	0	0	0	
Sentence level	P	365	0	0	75.25	100	85.88	75.26
	N	98	0	0	0	0	0	
	NEU	22	0	0	0	0	0	
Aspect level	P	562	0	0	77.73	100	87.47	77.73
	N	118	0	0	0	0	0	
	NEU	43	0	0	0	0	0	

Table 5. Proposed classification result using confusion matrices

Feature level		P	N	NEU	Precision (%)	Recall (%)	F-measure (%)	Accuracy
Title level	P	50	0	0	89.29	100	94.34	90.91
	N	3	10	0	100	76.92	86.95	
	NEU	3	0	0	0	0	0	
Document level	P	78	0	0	91.76	100	95.71	90.10
	N	0	10	0	76.92	100	86.96	
	NEU	7	3	3	100	23.08	37.50	
Sentence level	P	365	0	0	99.46	100	99.73	94.85
	N	0	95	3	82.61	96.94	89.20	
	NEU	2	20	0	0	0	0	
Aspect level	P	562	0	0	93.20	100	96.48	94.05
	N	0	118	0	98.33	100	99.16	
	NEU	41	2	0	0	0	0	

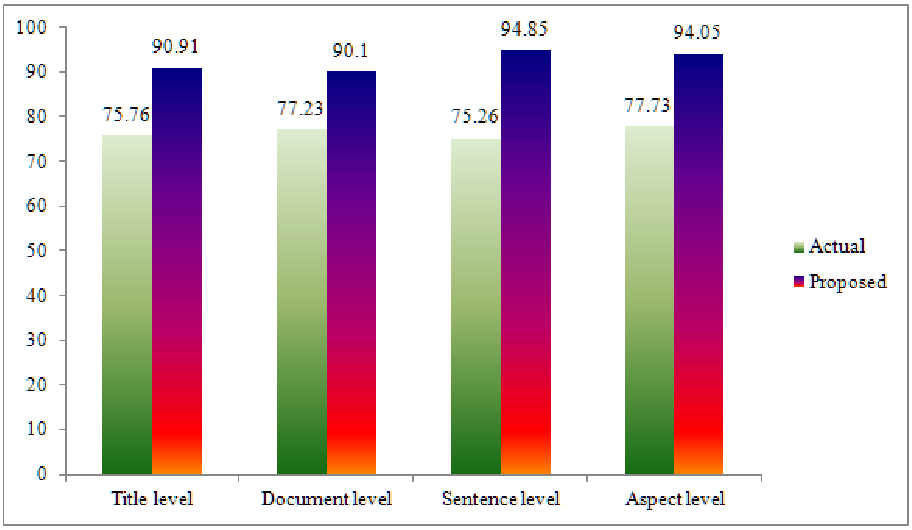

Figure 6. Comparison of accuracy between actual and proposed

In this system, the accuracy was obtained for the title level, document level, sentence level, and aspect level is 75.76%, 77.23%, 75.26%, and 77.73% respectively. The average accuracy of the actual system is 76.50%. The proposed system achieves the accuracy of 90.91%, 90.10%, 94.85%, and 94.05% for each class. The average accuracy of the proposed system is 92.48% with 69% precision and 66% recall.

CONCLUSION

Online social media contents are steadily growing day by day about the open and distance education. It's become more useful to the public for decision-making purpose, improving the quality of education, quality of service, creating policy, etc. In this research paper, a multi-aspect based opinion mining system is proposed for open and distance education to measure the satisfaction of the public . The proposed system is too general, it could be used in any fields like education, medical, engineering, science, etc. The system was implemented by applying various steps, namely, data collection process, data preprocessing, feature extraction, opinion detections, opinion visualization, classification, and evaluation. A highlight of this research is measuring the public opinion about IGNOU using online reviews, and by applying the Naïve Bayes sentiment Classification algorithm at the title level, document level, sentence level, and aspect level. The experimental results reveal that the proposed system achieved 92.48% accuracy than the actual system. In future research, the performance of the proposed system could be compared with Sentiment140, SentiStrength, and SentiWordNet, and also ranks the open and distance Universities based on aspects.

REFERENCES

A. Moreo, M. Romero, J.L. Castro, & J.M. Zurita (2012). Lexicon-based Comments-oriented News Sentiment Analyzer system, *Expert Systems with Applications*, 39, 9166 – 9180.

Andres Montoyo, Patricio Martinez-Barco, & Alexandra Balahur (2012). Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, *Decision Support Systems*, 53, 675-679.

Annelien van Rooyen (2015). Distance education accounting students' perceptions of social media integration. *Procedia - Social and Behavioral Sciences*, 176, 444-450.

Bahar BERBEROĞLU (2015). Open and distance education programs of Anadolu University since the establishment. *Procedia - Social and Behavioral Sciences*,174, 3358-3365.

Bing Liu (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.

Daniel E. O' Leary (2011). Blog mining-review and extensions: "From each according to his opinion", Decision *Support Systems*, 51, 821-830.

Deanne K. Bird, Katharine Haynes, Rob van den Honert, John McAneney & Wouter Poortinga (2014). Nuclear power in Australia: A comparative analysis of public opinion regarding climate change and the Fukushima disaster, *Energy Policy*, 65, 644-653.

Dr. Perihan Paksoy (2015). Use of Global Distance Education Program in Educational Goals of Airline. *Procedia - Social and Behavioral Sciences*, 177, 372-378.

Ebru Melek Koc (2012). Idiographic roles of cooperating teachers as mentors in pre-service distance teacher education. *Teachers and Teacher Education*, 28, 818-826.

Farhan Hassan, Saba Bashir & Usman Qamar (2014). TOM: Twitter opinion mining framework using hybrid classification Scheme, *Decision Support Systems*, 57, 245-257.

Fatma Kübra Çelena, Aygül Çelikb, Süleyman Sadi Seferoğlu (2013). Analysis of Teachers'Approaches to Distance Education. *Procedia - Social and Behavioral Sciences*, 83, 388-392.
Filiz Kanteka (2014). Distance Education in Nursing in Turkey. *Procedia - Social and Behavioral Sciences* 116, 639-643.
Galina Samigulina, Zarina Samigulina (2016). Intelligent System of Distance Education of Engineers, based on Modern Innovative Technologies. *Procedia - Social and Behavioral Sciences*, 228, 229-236.
Gao Guohong, Li Ning, Xiao Wenxian, Wan Wenlong (2012). The Study on the Development of Internet-based Distance Education and Problems. *Energy Procedia*, 17, 1362 – 1368
Haeng-Jin Jang, Jaemoon Sim, Yonnim Lee & Ohbyung Kwon (2013). Deep sentiment analysis: Mining the causality between personality-value-attitude for analyzing business ads in social media, *Expert Systems with Applications*, 40, 7492-7503.
Hieronymus J.M. Gijselaers, Paul A. Kirschner, Peter Verboon, Renate H.M. de Groot (2016). Sedentary behavior and not physical activity predicts study progress in distance education. *Learning and Individual Differences*, 49, 224–229.
Howard J. Hamilton (2012). Knowledge Discovery in Databases, Teaching Notes, University of Regina.
M.Banu Gündoğan, Gülsün Eby (2012). A green touch for the future of distance education. *Procedia - Social and Behavioral Sciences*, 55, 789-798.
Magdalini Eirinaki, Shamita Pisal & Japinder Singh (2012). Feature-based opinion mining and ranking, *Journal of Computer and System Science*, 78, 1175-1184.
Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009). The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, Volume 11, Issue 1
MeaningCloud™ (http://www.meaningcloud.com) have been used for Text Analytics purposes in *the development/testing/validation of this research/prototype/software.*
Nurmukhametov N, Temirova A, Bekzhanova T (2015). The Problems of Development of Distance Education in Kazakhstan. *Procedia - Social and Behavioral Sciences*,182, 15-19.
Oksan Bayulgen & Ekim Arbatli, (2013). Cold War redux in US – Russia relations? The effect of US media framing and public opinion of 2008 Russia – Gerogia war, Communist and Post – *Communist Studies*, 46, 513-527.
Pawel Sobkowicz, Michael Kaschesky & Guillaume Bouchard (2012). Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web, *Government Information Quarterly*, 29, 470-479.
Qiwei He, Cees A. W. Glas, Michael Kosinski, David J. Stillwell & Bernad P. Veldkamp (2014). Predicting self-monitoring skills using textual posts on Facebook, *Computers in Human Behavior*, 33, 69-78.
Safiullin Lenara, Fatkhiev Arturb, Saipullaev Ullubic, Bagautdinova Nailya (2014). Problems And Decision In The Field Of Distance Education. *Procedia - Social and Behavioral Sciences*, 131, 111-117.
Tahir Tavukcu, øbrahim Arap, Deniz Özcan (2011). General overview on distance education concept. *Procedia Social and Behavioral Sciences*, 15, 3999-4004.
Vasileios Kagklis, Anthi Karatrantou, Maria Tantoula, Chris T. Panagiotakopoulos, Vassilios S. Verykios (2015). A Learning Analytics Methodology for Detecting Sentiment in Student Fora: A Case Study in Distance Education. *European Journal of Open, Distance and e-Learning* – Vol. 18 / No. 2, 75-94.
Xiao Wang (2013). Applying the integrative model of behavioral prediction and attitude functions in the context of social media use while viewing mediated sports, *Computers in Human Behavior*, 29, 1538-1445.
Yingqi Tang (2013). Distance Education Librarians in the United States: A Study of Job Announcements. *The Journal of Academic Librarianship*, 39, 500-505.