International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-3, January 2020

Sentiment Mining of Product Opinion Data

Vikas Thada, Utpal Shrivastava, Ruchi

Abstract: Sentiment Analysis has gotten a focal point of consideration for the retrieving of web information and text corpora. At the point when we talk about sentiment analysis, it investigates the people's criticism or conclusion or audits in the wake of removing those reviews from various stages. While, when we find out about Sentiment Analysis, it recognizes the mining or analysis, which are communicated through content and afterward we investigate it. The fundamental goal of Sentiment Analysis is to reason feelings and ideas. In this paper the research work has done sentiment analysis on Amazon Reviews. The data set used is Amazon Reviews on Unlocked Mobile phones dataset of 413840 records. There are number of columns for the dataset like Product Name, Price, Rating, Brand Review text and the count of people for whom the review was helpful. But the research work has used on the Rating and Reviews columns. The research work has applied bag of words, tf-idf and n-grams techniques. It was found that n-gram with bag of words approach gave a maximum testing accuracy of 97% and AUCROC score of .967. Using deep learning approaches and their comparing with used ones is left as future work.

Keywords: Sentiment, Analysis, Opinion, mining, review, bag of words, n-gram.

I. INTRODUCTION

Sentiment Analysis has gotten a focal point of consideration for the retrieving of web information and text corpora. At the point when we talk about sentiment analysis, it investigates the people's criticism or conclusion or audits in the wake of removing those reviews from various stages. While, when we find out about Sentiment Analysis, it recognizes the mining or analysis, which are communicated through content and afterward we investigate it. The fundamental goal of Sentiment Analysis is to reason feelings and ideas . The sentiments are given by person or company or any organization and are about any physical thing or product or different features of a product such as mobile,tv,vehicle etc. Based on what kind of sentiments are given about any product or features the sentiments can be positive, negative or neutral that is known as orientation or polarity of the sentiment[1,2].This is shown in figure 1.

For example, “Picture quality of the mobile is great”. The user is someone who has used the mobile and reviewing it that is expressing his/her opinion/sentiment about the product’s features and it is positive because of “great”. Thus from the review we determined the semantic orientation and concluded that orientation/polarity is positive [1].

Revised Manuscript Received on January 30, 2020.
***** Correspondence Author
Dr. Vikas Thada*, Department of Computer Science & Engineering Amity University Gurgaon.
Mr Utpal Shrivastava, Department of Computer Science & Engineering Amity University Gurgaon.
Ms Ruchi, Department of Computer Science & Engineering Amity University Gurgaon.

We can say that assurance of semantic direction is an errand of finishing up whether a sentence or review has polarity as shown in figure 1 [1].

beliefs. Multiple opinions may be present in a subjective sentence. An example of objective sentence: There was a reunion of batch 2008 after 10 long years and it was by some dramatic coincidence. An example of subjective sentence: *The service of this restaurant is wonderful. I always enjoy their hospitality.*

In aspect/feature level the aim is to looking for features of objects/product/entity that have been reviewed by individual person or company and classifying into positive, neutral or negative. Similar words based on same lemma or synonyms are combined together and collection of reviews prepared [2].

This is approach usually employed when performing sentiment analysis of reviews of mobile phones, restaurants, cars, jackets or any other physical thing.

Sentiment analysis contains a wide scope of utilizations, beginning from following the clients' sentiments concerning item or concerning political applicants as communicated in on-line gatherings and for customer relationship management [3]. Sentiment can be gotten through different ways about numerous zones like an entity/product or service. Sentiment Mining might be significant in various manners. For example, it judges and tracks the achievement pace of propelling another thing or a business drive, centers pervasiveness of organizations and things with its forms. It additionally gives us information about socioeconomics of explicit qualities are being preferred or loathed.

The paper is organized as follows: section 2 related work reviews some of the best work carried out in the area of sentiment analysis, section 3 carries out the actual research work done along with analysis, section 4 concludes with some directions for future work.

II. RELATED WORK

Plenty of work has been done on the topic of semantic analysis. This section survey some of the related work.

Pang et al[4] worked on movie review data by utilizing the machine learning techniques involving SVM,SME and Naïve Bayes with n-grams (1 and 2). They concluded that SVM resulted in better accuracy and Naïve Bayes as poor. Hu et. al using amazon and cnet dataset for their work . After extracting adjectives they made use of WordNet [7] to check for negative or positive sentiment . Their research work made use of techniques that were lexicon based for finding adjective orientation. They also used NLP based parser that employing Part-Of-Speech (POS) tagging for assigning tags to individual tokens. Ohana et.al [8] shows the results of utilizing the Senti Word Net to the issue of opinion characterization of information got from film audits [8].

Their examination assesses the work of SentiWordNet to the content level characterization by methods for the dataset displayed in [9].
Their methodology includes summarizing positive and negative term scores to assess feeling direction, and change is appeared by bringing a dataset of related choices exploitation SentiWordNet, and rehearsed to an AI classifier. Prabowo et al. [9] has common basic based for the most part game plan, AI and machine learning techniques. This technique is confirmed on motion picture appraisals, creation evaluations and MySpace comments. Ghag et al [10] likewise performed opinion mining by utilizing diverse classifier. Systems thought about based on language reliance, Usage of dictionary and preparing sets. Significant difficulties incorporate taking care of refutation and language speculation. In 2014, Gupte et al. [11] performed a comparative analysis of various algorithms like Score aggregation, text refinement, text classification and extraction. Their research showed great affinity for Random Forest classifier because of execution, clear comprehension, and gradual results although the preparation time was high but of high accuracy. They also preferred NB on account of its shorter memory utilization and preparing time. There is a larger than usual corpus of printed contemplates portraying the job of Sentiment Analysis (SA). Bhadane et.al[12] arranged a framework. The Analysis is of two stages: First stage identify aspect and second stage performs classification of sentiments They completed a game plan of frameworks for highlight extremity documentation and characterization of innovation assessment by methods for learning (SVM) imparted to territory word references. Their test exhibited that the prescribed strategies have achieved around "78%" accuracy and are very promising in playing out their tasks[11]. Mandal and et.al [13] introduced proposed text classification algorithm based on Lexicon . Their method was based on dictionary that used emotions lexis in looking for word within sentences to check the polrariy. Their work worked well for on-line reviews. Pasarate et.al [14] used movie reviews and did parsing using Stanford parser . They made use of different extraction techniques. TWSC Method, CM Model, AW Scheme, IEDR. Out of all 4 approaches the IEDR turned out to be a winner. This method shows highlight extraction for the improvement of execution upgrades in the correlation of other elective systems utilized for the opinion mining and analysis. Because of the complex phonetic qualities, feelings analysis is finished at various phases of substance. Suresh and et. al [15] proposed technique utilizing RatingSystem.com informational index, and the exploratory results indicate results that the introduced highlight choice system is promising. In their work they gave significance on include determination for feeling investigation utilizing choice trees. There are different difficulties engaged with Sentiment assessment. To improve the web shopping data and interface with customers through the intensity of the thing evaluations, studies, customer Q&A and individual to individual correspondence, Rating System.com is assisting associations around the world. They used two procedures named as Proposed Feature Selection dependent on Decision Trees and LV Quantization. The PCA and the proposed highlight determination technique was utilized to diminish the highlights. Gullible Bayes with LV Quantization has high exactness. They uncovered the arrangement precision got from LVQuantization and contrasted and NB classifier, CART. The characterization accuracy overcame Naïve Bayes with LVQ is superior to anything Naïve Bayes with PCA. They additionally display the RMSE. It very well may be seen that the exactness and survey is down for these classifiers. In any case, it was seen that the exactness for positive conclusions was moderately low. This wonder was seen with LVQ just as with NB classifier [16]. Rana et.al[17] using dataset of movie reviews made a Proportional Examination of Emotional Location By methods for SVM and Naïve Bayes Techniques. They recognize the assessment of individuals. Naive Bayes utilizing manufactured word approach and direct SVM came about to give best precision. They likewise indicated Polarity of various words in diagram. In further they need to investigate this information with various items and territories. Now a days individuals want to purchase online items, so along these lines exactness pace of items can undoubtedly be recognized [17].
M.Shaheen et al.[18] et al performed a variety of techniques on mobile review data with random forest and deep learning techniques like CNN and LSTM that provided better accuracy.

III. METHODS AND MATERIALS

In this paper the research work has done sentiment analysis on Amazon Reviews. The data set used is Amazon Reviews on Unlocked_Mobile phones dataset of 413840 records. There are number of columns for the dataset like Product Name, Brand, Price, Rating, Review text and the numberof people who found the review helpful. But the research work have used on the Rating and Reviews columns. The research work was carried out using Anaconda 4 framework in jupyter notebook. The steps are as follows:

I. Fetching of datasets
II. Cleaning and Normalization of data sets
III. Removing of neutral values and encoding of ratings.
IV. Extraction of features (bag of words approach) using count vectorizer
 tf-idf and tf-idf with n-gram (3 different approach)
V. Fitting and transforming in each approach
VI. Classification using Logistic Regression
VII. Calculating AUC_ROC score
VIII. Comparing 3 approaches

IV. EXPERIMENTATION & RESULTS

The task started by cleaning the dataset (removing null values) followed by converting the ratings into just two values so classification becomes binary only. Removing null values from the document leaves dataset with 334335 records. The dataset was prepared for binary classification removing neutral values, encoding values greater than 3 to 1 (positive reviews) and less than three as 0(negative reviews). This resulted in dataset of 308277 values. The figure 2 below shows graphical representation among reviews and ratings:

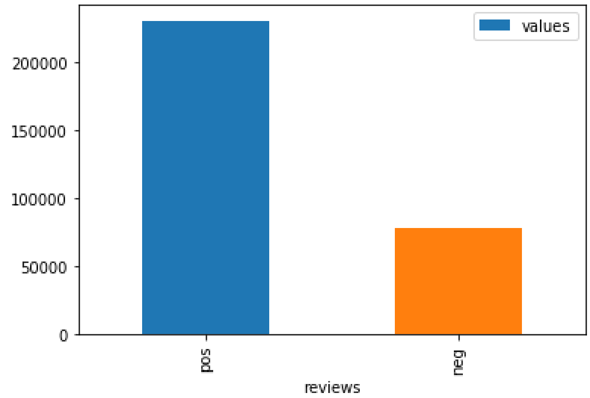

Fig 2: Bar graph of two polarity(positive and negative)

This dataset of 308277 was split in training set and testing set in the ratio of 80:20. We used only reviews and their rating columns for the training. The ratings serve as targets and reviews as features. The split gives training set of size 246621 records and testing set of 61656 records.

Now before we feed the training set for training purpose we need to convert reviews into a numeric representation that training algorithms can make use of. The approach used in the research is bag-of-words approach. It is frequently employed in machine learning for text representation and shown in figure 3.

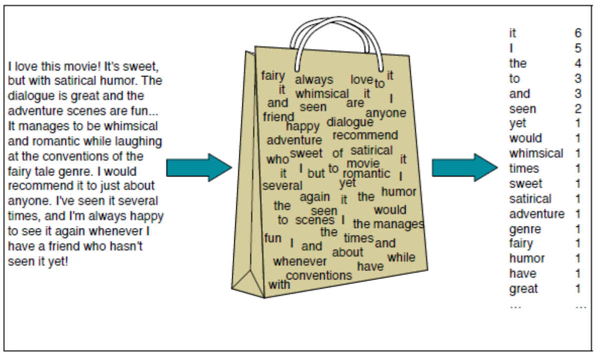

Fig 3: Bag of Words Example [19]

In this approach the order of words and their position is not considered instead only the frequency of the words in the document is kept.

The CountVectorizer from scikit-learn is ideal candidate to do this job by applying bag-of-words approach to collection of text documents and returning a matrix of token counts. Fitting the CountVectorizer with training data consists of the tokenization of the trained data and building of the vocabulary. Fitting tokenizes each document by finding all sequences of characters of at least two letters or numbers separated by word boundaries. Then it is converted to lowercase and token vocabulary is created. Some of the tokens are shown below:

['00', '4and', 'adhesive', 'artículos', 'biometrics', 'caras', 'completar', 'dalete', 'display2', 'envia', 'feew', 'getvroot', 'hideous', 'instuctions', 'laos', 'mascara', 'myphoneexplorer', 'opamp', 'phoneor', 'promises', 'referencial', 'sailor', 'sipper', 'stomach', 'thanksta', 'ultimately', 'volts', 'yesterday']

This gives us a feeling as how our vocabulary looks like. The total features we got after fitting the train data is 54467. By looking just few number of tokens we can see that its not pretty clean as it contains unwanted words with no English meaning as well as misspelled words too.

The vector object thus used for transforming the trained data to document term matrix, giving us the bag-of-word representation of trained data. In the matrix the row represents document and column represent training vocabulary word. Thus each intersection of row and column represent count of particular word in that document. Further it is sparse matrix as most of the entries are zero. On example is shown in figure 2.

Document 1	Love the life you have
Document 2	Love you my zindagi
Document 3	Live and let live
Document 4	Live life with love

Feature names: ['and', 'have', 'let', 'life', 'live', 'love', 'my', 'the', 'with', 'you', 'zindagi']

Doc/Feat.name	and	have	let	life	live	love	my	the	with	you	zindagi
Doc 1	0	1	0	1	0	1	0	1	0	1	0
Doc 2	0	0	0	0	0	1	1	0	0	1	1
Doc 3	1	0	1	0	2	0	0	0	0	0	0
Doc 4	0	0	0	1	1	1	0	0	1	0	0

Fig 4: Example of document term matrix

Lets call this matrix as our feature matrix. To train our model Logistic Regression technique was used and this feature matrix along with target ratings values were passed. To test our model test data were used and performance was measured in terms of ROC_AUC curve. But before that test data was also transformed similar to transformation of training data. As our vectorizer was already fitted ,only transformation was applied on test data.

Looking at our AUC score,we see we achieve a score of about 0.927. This is shown in the figure 5.

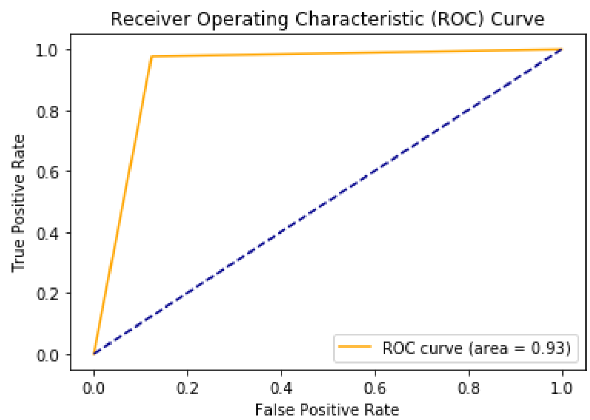

Fig 5: ROC_AUC curve has score of 0.927

If we look at the contribution from the coefficients that are responsible for predicting positive or negative sentiments. The top ten features for positive sentiments are

['excelent' 'excelente' 'excellent' 'exelente' 'loving' 'loves' 'perfecto' 'amazing' 'lovely' 'love']

Fig 6: Top 50 features for positive sentiments

And top ten features for negative sentiment are:

['worst' 'mony' 'false' 'worthless' 'junk' 'garbage' 'useless' 'lies' 'horrible' 'nope']

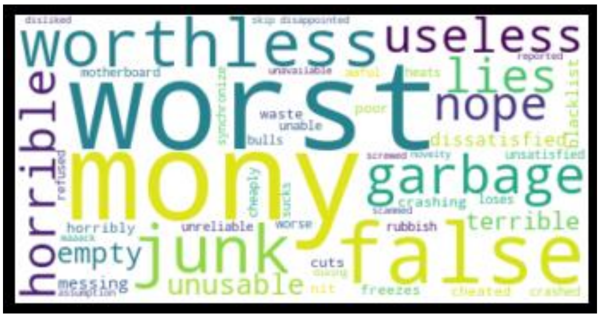

Fig 7: Top 50 features for negative sentiments

4.1 Second Approach

The second approach we have used for sentiment analysis is to use tf-idf (term frequency-inverse document frequency). The features will be rescaled using tf-idf technique. Term frequency-inverse document frequency, allows us to weight terms based on how important they are to a document. High weight is given to terms that appear often in a document, but don't appear often in the corpus. Features with low tf–idf are either commonly used across all documents or rarely used and only occur in long documents. Features with high tf–idf are frequently used within specific documents, but rarely used across all documents.

Like in the previous approach of using CountVectorizer, features returned by tf-idf vectorizer are same but here we want to reduce the number of features that might help improve our model's performance. The performance was improved by neglecting words that appear in very few documents say 5,6 or any number. This helps us remove some words

that might appear in only a few and are unlikely to be useful predictors. For example, comparing the length of the features in both approaches, this approach has just 18621 features against 54467 features in first approach. This has resulted in approximately 66% reduction in number of features that were not useful predictors.

Next, when we transform and fit our training data. Once model is ready predictions are done on transformed test data . We can see we, again, get an AUC of about 0.927. No improvement in AUC score, but we were able to get the same score using far fewer features.

If we look at the contribution from the coefficients that are responsible for predicting positive or negative sentiments. The top ten features for positive sentiments are

['love' 'great' 'excellent' 'perfect' 'amazing' 'awesome' 'perfectly' 'easy' 'loves' 'best']

Fig 8: Top 50 features for positive sentiments

And top ten features for negative sentiment are:

['not' 'worst' 'useless' 'terrible' 'disappointed' 'return' 'waste' 'horrible' 'poor' 'doesn']

Fig 9: Top 50 features for negative sentiments

4.2 Improving accuracy with n-gram model

The BOW approach used earlier has a problem that order of word does not matter. This has a direct impact on the performance of the model. Take an example of two different Sentences: *'Not bad mobile'* and *'mobile is bad'*. The first one is positive sentiment and other one isNegative sentiment. But for both sentences our model predicts negative sentiment. . To put some context into searching we can make use of n-grams (sequences of 2 or more features) . We can have two words(bigram), three words(trigram) or more words combination as our feature.

For example, bigrams, is a sequence of two words, could give us features such as

Not bad versus is bad. And trigrams,which give us triplets of adjacent words, could give us features

such as not much issue. The count vectorizer technique was used to train again so that it returns features of 1-gram and 2-gram . By adding bigrams to the model the AUC score obtained was 0.967 and its an improvement over .927 got earlier without n-gram figure 10 and table 1.

If we look at the contribution from the coefficients that are responsible for predicting positive or negative sentiments. The top ten features for positive sentiments are

['not bad' 'excelent' 'excelente' 'excellent' 'perfect' 'no problems' 'exelente' 'awesome' 'great' 'no issues']

And top ten features for negative sentiment are:

['no good' 'junk' 'worst' 'not good' 'not happy' 'horrible' 'garbage' 'terrible' 'not satisfied' 'not worth']

Keep in mind that, although n-grams can be powerful in capturing meaning, longer sequences can cause an explosion of the number of features. Just by adding bigrams, the number of features we have, increased to almost 209,766. .

If we take a look at what features our model connected with negative reviews, we can see that we now have bigrams such as no good and not happy, while for positive reviews we have not bad and no problems. If we again try to predict *'not bad'*, and *'is bad'* we can see that our newest model now correctly identifies them as positive and negative reviews respectively.

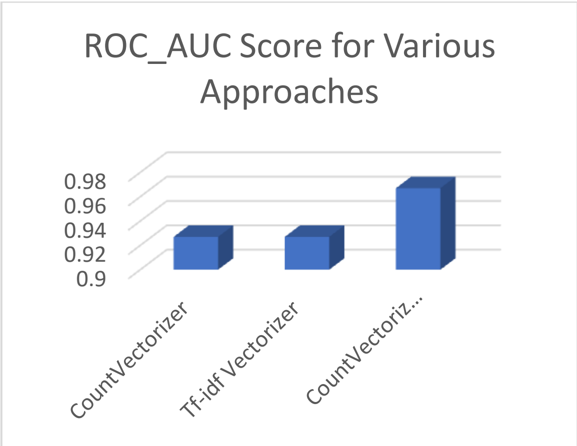

Fig 10: ROC_AUC Score for various approaches

Table 1: ROC_AUC Score

ROC_AUC Score	
CountVectorizer	0.927
Tf-idf Vectorizer	0.927
CountVectorizer with ngram	0.967

V. CONCLUSION

The research work has carried sentiment analysis of product (mobile phones) review using bag of words approach employing count vectorizer, tf-idf vectorizer with n-gram model. The aim was to get the best AUC_ROC score with testing data using varied techniques. The dataset used was unlocked mobile phones from amazon. The performance was measured in terms of AUC_ROC curve. It is concluded that tf-idf with n-gram model gave AUC score of 0.967 as shown in table 1 and figure 10 and was best among the three. The comparison with other techniques such as deep learning models for sentiment classification is left as future work.

REFERENCES

1. R. Sharma, S. Nigam, and R. Jain, "Supervised Opinion Mining Techniques : A Survey," vol. 3, no. 8, pp. 737–742, 2013.
2. https://monkeylearn.com/sentiment-analysis/
   [2] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Eng. J., vol. 5, no. 4, pp. 1093–1113, 2014.
3. A. Esuli and F. Sebastiani, "SENTIWORDNET: A high-coverage lexical resource for opinion mining," Evaluation, pp. 1–26, 2007.
4. B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," 2004.
5. M. Hu and B. Liu, "Mining and summarizing customer reviews," Proc. 2004 ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '04, p. 168, 2004.
6. M. Hu and B. Liu, "Mining Opinion Features in Customer Reviews," 19th Natl. Conf. Artifical Intell., pp. 755–760, 2004.
7. G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller, "Introduction to wordnet: An on-line lexical database," Int. J. Lexicogr., vol. 3, no. 4, pp. 235–244, 1990.
8. B. Ohana and B. Tierney, "Sentiment classification of reviews using SentiWordNet," Sch. Comput. 9th. IT T Conf., p. 13, 2009.
9. R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach," J. Informetr., vol. 3, no. 2, pp. 143–157, Apr. 2009.
10. K. Ghag and K. Shah, "Comparative analysis of the techniques for Sentiment Analysis," Int. Conf. Adv. Technol. Eng., no. 124, pp. 1–7, 2013.
11. A. Gupte, S. Joshi, P. Gadgul, and A. Kadam, "Comparative Study of Classification Algorithms used in Sentiment Analysis," Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 5, pp. 6261–6264, 2014.
12. C. Bhadane, H. Dalal, and H. Doshi, "Sentiment analysis: Measuring opinions," Procedia Comput. Sci., vol. 45, no. C, pp. 808–814, 2015.
13. E. K. Fi et al., "a Lexicon-Based Text Classification," 2016.
14. S. Pasarate and R. Shedge, "Comparative study of feature extraction techniques used in sentiment analysis," 2016 Int. Conf. Innov. Challenges Cyber Secur., no. Iciccs, pp. 182–186, 2016.
15. I. Journal, "Sentiment Classi fi cation using Decision Tree Based Feature Selection," no. January 2016, 2017.
16. A. J. Singh, "Sentiment Analysis : A Comparative Study of Supervised Machine Learning Algorithms Using Rapid miner," vol. 5, no. Xi, pp. 80–89, 2017.
17. S. Rana and A. Singh, "Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques," in 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), 2016, pp. 106–111.
18. M.Shaheen, "Sentiment Analysis on Mobile Phone Reviews Using Supervised Learning Techniques",
    I.J. Modern Education and Computer Science, 2019, 7, 32-43
19. D.Jurafsky,H.Martin "Speech & Language Processing", Wiley Publications

AUTHORS PROFILE

Dr. Vikas Thada, has doctoral degree in Computer Science & Engineering from Dr. K.N.M University Newai, Jaipur, India, Master's Degree in Computer Science & Engineering from Rajasthan Technical University Kota, Rajasthan, India and Bachelor's Degree also in Computer Science & Engineering from Engineering College Kota, Rajasthan, India. He is currently serving as Associate Professor in the Department of Computer Science & Engineering, Amity School of Engineering & Technology, Amity University Haryana. He has more than 19 years of teaching experience with around 8 years of research experience. He has many publications in international journals and is author of number of books on programming, data structures etc. His research interests are genetic algorithm, cryptography and network security, machine learning and natural language processing.

Mr. Utpal Shrivastava, has M. Tech in computer science and engineering form GJUS&T, Haryana. He is pursuing his doctoral degree from Banasthali University Jaipur. He has 12 years of teaching experience and currently serving as assistant professor at Amity University, Haryana in the department of computer science and engineering. His research interest is in machine learning, genetic algorithm, image processing and security attacks.

Ms. Ruchi Kamra, has obtained M.Tech in Computer Science from GGSIPU,Delhi & B.Tech in Computer Science from Kurukshetra University, Kurukshetra. She is pursuing PhD from CCS Univertsity,Meerut. She is presently serving as Assistant Professor in Computer Science and engineering department at Amity University, Haryana and having experience of 9 years in teaching. Her areas of interest are Software Engineering, Data Mining and Big Data Analysis.