# Strategic Information Assessment: Polarity, Imbalance Data

**Naima Sabir*†, Dr. Muhammad Awais ††, , Zubair Nabi††† Muhammad Yahya Saeed††† †**

Government College University, Faisalabad, Punjab, Pakistan

**Abstract**
Digital data available on internet becoming a major source of all types of information. There is a strategic value of information on various topics. The strategic value of data is dependent on the nature, quality, aspects, origination etc. The assessment of data has different meaning for every stakeholder. This research is based on the major assessment of data in form of balance and imbalance of data, uncertain and certain data. These two things are related to polarity of data. The micro and macro level polarity is dependent over the certain and balance of data this research maps the classification of data over these parameters.

***Keywords:***
*Digital Data, Data assessment, Imbalance Data, Polarity.*

## 1. Introduction

From Understudy criticism can feature different issues that understudies may have with the address. One case of this is the point at which an understudy does not see some portion of an address or a particular case. Another illustration is the too quick or too moderate pace of educator preparing. Input is typically gathered toward the finish of the square, however it is all the more positively taken progressively.

### 1.1 Polarity Recognition

This phase presents two examinations to decide extremity from understudy input. The past writing identified with the phases of notion examination in the discovery of extremity is considered. This phase closes with the best model found to distinguish extremity from understudies ' criticism continuously. At long last, an outline of the phase and materials is given [1].

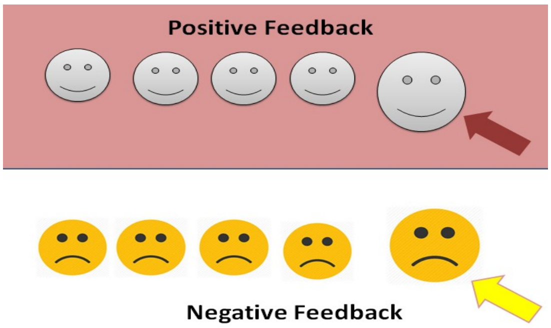


**Figure 1** Explanation of Data Polarity

### 1.2 Extremity recognition

The point of this investigation is to decide a supposition examination demonstrate adjusted for the instructive field, specifically for constant criticism in the classroom. There are four essential strides to distinguish examples of estimation examination utilizing the strategy for machine learning for deciding the extremity: information preprocessing, include choice, utilization of machine learning techniques and assessment of results. Past examinations identified with them are talked about in the accompanying subsections [2].

### 1.3 Preprocessing for Polarity

Preprocessing is a critical advance in getting ready information for examination. There are numerous regular strategies for pre-preparing in the discovery of extremity, of which the most well-known are: tokenization, changing over content to Lower or capitalized, evacuating

Manuscript received December 5, 2025
Manuscript revised December 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.12.16

accentuation, expelling numbers, evacuating copy letters, evacuating stop words, stemming and disavowal [3]. A portion of these preprocessing strategies are found in a past key examination think about, as portrayed beneath:

1. Tokenization: part sentences into words, expressions, and images. Tokenization was utilized in Mouthami.
2. Concealed content in Lower or capitalized: change over letters to upper or Lower case to coordinate them with different events in the preparation information. The word in capitals now and again talks about forceful feelings [4]. One case showing this is "I fizzled the exam"; the word" fizzled" can demonstrate solid negative or baffling feelings.
3. Expel accentuation denotes: the evacuation of accentuation marks, for example, question marks and shouts marks, has been connected by numerous scientists, for instance [5]. Accentuation for the most part does not hold any wistfulness, notwithstanding, there are circumstances when it can pass on the intensity of nostalgia; for instance, in " I passed!!!an shout point can demonstrate solid positive or euphoric feelings. What's more, question marks can be befuddling.
4. Erase numbers: erasing numbers is regular in past examinations; for the most part one case of numbers that can speak to inclination is "awesome", which can be composed in talk dialect as "gr8". Albeit a few numbers may reflect conclusion, they are wiped out much of the time.
5. Expelling copy letters: spelling mistakes can be redressed by evacuating additional letters [6]. A few scientists have safeguarded two letters and evacuated the additional rehashed letters [64, 166, 181]. Different specialists, for example, Agarwal et al. [7], has safeguarded three letters and erased any extra.
6. Evacuate stop words: There is no steady arrangement of stop words that must be expelled.
7. Rising: Returning the word to the essential frame; for instance "learning" will be "learn". Cases of analysts who have utilized undifferentiated cells.
8. Spell check: spell check for incorrectly spelled words is additionally an essential preprocessing strategy because of misclassification or overlooking incorrectly spelled words. Cell phones and long range interpersonal communication locales currently offer spelling adjustments, committing spelling errors less normal.
9. Invalidation: refutation changes the extremity of a word, i.e. positive to negative or negative to positive. A few cases of deniers are 'no', 'no', 'nobody', 'never' and 'nothing' [8]. In Song et al. [9] for the recognizable proof of the sentences of dissent was made a calculation in view of guidelines. Container and Lee [10] additionally found nullification in the sentence. A few cases of deniers are given in table 1. There are a few troubles with recognizing refutation, for example,

* The place of refutation may influence the sentence.
* The refutation of sentences when the nullification shows up before the words; for instance, "supernatural occurrence": "it isn't astonishing that everybody cherishes him" [11].

**Table: 1 Negative Examples**

| Negative | Example |
|---|---|
| **Nobody** | **Nobody can understand the leacture of CS.** |
| Nothing | Just another subject, nothing spectacular |
| None | None of them become leader . |

Notwithstanding the preprocessing techniques depicted above, there are distinctive kinds of preprocessing strategies utilized relying upon the information source. For instance, if the information were gathered from Twitter or online life, distinctive pre-handling strategies, for example, supplanting hashtags and unique Twitter images, for example, emojis, would be required to gather the information on paper. Probably the most widely recognized Twitter-particular information preprocessing methods are: disposal, the distinguishing proof of emojis [12].

The majority of the above pretreatment techniques were utilized in many past examinations in the field of instruction [13]. Spell checking is a typical strategy utilized in the instructive space and worked. This technique is typically utilized with information gathered from interpersonal organizations that contain the visit dialect. Invalidation is another normal preparing system that influences extremity [14]. In training, it is found in concentrates, for example.

### 1.4 Highlights of the Polarity Phase

The highlights give a more precise examination of temperaments and a nitty gritty speculation of the outcomes [15]. The most widely recognized signs are N-gram [16] and POS (grammatical feature) marks [17]. Also, there are less basic information source-related capacities (i.e. Twitter-related or area particular capacities). They are clarified in the subsections underneath, which likewise incorporate a review of pertinent investigations utilizing these highlights.

## 2. Material and Methods

This phase gives a definite portrayal of the exploration procedure of this paper. Area 2.1 presents about problem statement. Prior to giving a point by point strategy, the issues, goals and targets of the investigation are discussed in areas 2.2. About information assessment we concluded in phase 2.3. Emotion understanding is discussed in 2.4.

### 2.1 Problem Statement

Past writing recommends that manual input investigation is tedious and upsetting. In this postulation, this issue is unraveled by proposing a framework that can naturally investigate understudies ' criticism and present it to the educator continuously[18,19,20]. To make a framework, investigation is utilized to examine criticism. Past writing does not uncover assessment examination models reasonable for investigating understudy input. Thusly, this examination investigates different properties of notion examination models to distinguish new models. There is additionally an absence of investigation into imagining this sort of information, so this examination takes a gander at various methods for showing the information seriously.

### 2.2 Aims and Objectives

This postulation is given to the making of an arrangement of investigation of understudies ' input and displays it to the instructor continuously. The framework comprises of characterizing assessment examination models that give ideal outcomes and making perceptions that are important and helpful to the speaker[21,22]. With a specific end goal to accomplish the above destinations, the accompanying targets have been recognized:

1. Distinguish the most suitable models for anticipating extremity and feeling, which incorporate deciding the best blend of preprocessing procedures, capacities, and machine learning techniques.

2. Investigate diverse methods for displaying results and make perceptions that make it simpler to comprehend the aftereffects of opinion examination models.
3. Assess the framework in genuine addresses.

### 2.3 Information Assessment

In the past phase, we exhibited our extremity location tests, and in this Chapter we depend on the aftereffects of these analyses to contemplate feeling recognition. We audit past writing identified with feeling recognition and examine the investigation of feelings. In this phase, we show two examinations: the investigation of different capacities in segment[23,24]. This phase finishes up with ideal models for identifying feelings in understudies ' criticism progressively. At last, a rundown of the Chapter and the material introduced is given.

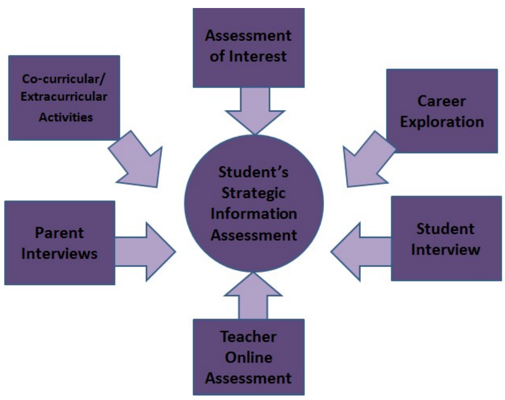


**Figure. 2** Explaining Information Assessment

### 2.4 Emotion Understanding

Feelings are imperative in the learning procedure [25,26]. Positive feelings can expand understudies ' enthusiasm for learning, increment inclusion in classes and persuade understudies [27]. What's more, upbeat understudies have a tendency to be more spurred to accomplish their learning objectives [28,29]. Then again, feelings, for example, outrage, nervousness, and misery negatively affect understudies and can divert them from learning.

**Table: 2** Evaluation of the Work

| Classes | Detail of Work |
|---|---|
| Accuracy | T P+T N+FP+FN |
| Exactness | T P |
| Recall | T P+FN |
| F score | F=2 Precision×Recall . × Precision+Recall |

## 3. Result And Discussion

With respect to the preprocessing contrast in connection to POS labels, which are given in table 2, the execution of the models are differed. KSD-b/n, RB, half W/O and me-watt display/O. showed itself with the best execution with the most extreme level of pre-treatment. We see that the preprocessing level of P4 has decreased the execution of most models when utilizing POS labeling capacities, inferring that evacuating stop words with POS labeling expels profitable data[30].

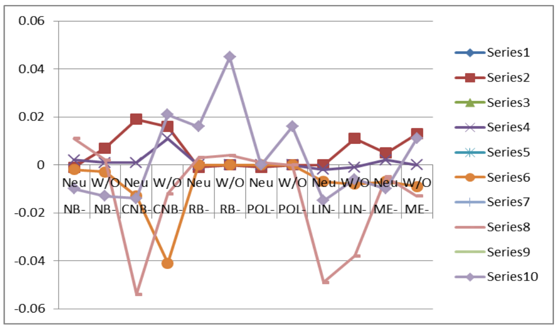


**Figure. 3** Performance between preprocessing levels for POS-labeling

To get an entire photo o f the preprocessing levels, we ascertained the execution contrast for all models between the preprocessing levels. Normal qualities for the precision of all models for each level were taken; they are introduced in fig 3. We found that the best pretreatment level is P6, aside from NB-W/O, CNB-Neu, LIN and ME demonstrates. With respect to the NB-W/O and CNB-Neu models, the most elevated execution was gotten at the P4 level[31,32], which shows that the poles and invalidations diminish the execution of these models. Like the h-grams handling results contrast, the Line demonstrated the best development with the least level of preparing. What's more, like the consequences of POS labels preprocessing, LIN-W/O demonstrated the best execution on P2. Likewise, the ME-W/O demonstrated the most astounding execution at the P2 level, while the ME-Neu demonstrated its best execution at the P3 level[33].

The primary inquiry investigated if understudies partook in giving input or not. The outcomes demonstrate that one understudy replied 'No' to this inquiry and two understudies left the appropriate response clear. One example t-tests were additionally utilized for each inquiry. One example t-tests are utilized to contrast the mean of an example with a specific esteem. The normal mean estimation of the three inquiries was computed which we name understudy question normal. The understudy question normal was utilized to investigate the general mean over the three inquiries [34]. To build up if the understudies loved or loathed the utilization of the framework, the methods for the single inquiries and the understudy question normal were contrasted and the center estimation of the scale, which shows an unbiased position. The aftereffects of the understudy one example t-tests are introduced in Table. The main contrast that was not noteworthy was in Q4 (i.e. the teacher changes the style of instructing). The understudy question normal mean esteem was 2.63(t(79)=-2266,

**Table: 3** *Question Detail*

| Sample | P-Value | Mean Value |
|---|---|---|
| Q2 | 0.024 | 2.69 |
| Q3 | 0.031 | 2.57 |
| Q4 | 0.057 | 2.68 |
| Composite Average | 0.024 | 2.63 |

p-value= 0.026), which demonstrates that the outcomes are fundamentally lower than the center esteem

1. Therefore, this recommends the understudies have a generally negative impression of the framework.

## 4. Conclusion

Concentrates identified with the identification of feelings were likewise considered. Writing has demonstrated that there is little research identified with the location of feelings through content. Past investigations have underscored the significance of feelings in learning. The writing brought up that there is no exploration that would ponder the location of feelings in criticism from understudies in the classroom[35,36]. Additionally, to the extent we know, there has been no examination on content information utilizing machine learning in the instructive setting, except for one investigation of the Chinese content. The primary commitment of this paper was to explore the extremity of discovery in understudies ' criticism. This commitment is set out in phase 2 . Two examinations were directed: the investigation of n-grams and POS-labels and the investigation of different highlights.

The principal test included the investigation of different levels of pre-treatment with the most widely recognized highlights, for example, n-grams and POS labels [37,38] and additionally the most famous machine learning strategies. Furthermore, a nonpartisan class was likewise examined in this trial[39]. One of the commitments to this examination was the investigation of the impact of pretreatment on extremity. In this manner, in the wake of contemplating 6 distinct levels of pre-treatment, we found that preprocessing builds the execution of the model. The second commitment was to think about the normal highlights in the writing (n-grams and POS-labels). Our outcomes demonstrated that N-grams brought about preferred execution over POS labels for our analysis. Another commitment was to distinguish the best machine learning calculations for identifying extremity in understudy criticism. Therefore, we found that LIN, NB, and CNB was the most astounding. Also, the nonpartisan class in our application was viewed as imperative, and the CNB classifier functioned admirably with unbiased class models[40].

## References


[1] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. Sentiment analysis of twitter data. In Proceedings of the Workshop on Languages in Social Media, pages 30–38, 2011.

[2] Asha Gowda Karegowda, AS Manjunath, and MA Jayaram. Comparative study of attribute selection using gain ratio and correlation based feature selection. International Journal of Information Technology and Knowledge Management, 2(2):271–277, 2010

[3] Amanda L Andrei. Development and evaluation of tagalog linguistic inquiry and word count (liwc) dictionaries for negative and positive emotion. Technical report, The Mitre Corporation, 2014. URL http://www.mitre.org/sites/default/files/publications/pr_14-3858-development-evaluation-of-tagalog-linguistic-inquiry.pdf.

[4] Amandeep Dhir, Khalid Buragga, and Abeer A Boreqqah. Tweeters on campus: Twitter a learning tool in classroom? Journal of Universal Computer Science, 19(5): 672–691, 2013.

[5] Arthur Graesser, Patrick Chipman, Brandon King, Bethany McDaniel, and Sidney D'Mello. Emotions and learning with auto tutor. Frontiers in artificial intelligence and application, 158:117–139, 2007.

[6] Andrew F. Hayes and Klaus Krippendorff. Answering the call for a standard reliability measure for coding data. Communication Methods and Measures, 1(1):77–89, 2007.

[7] Albert Bifet and Eibe Frank. Sentiment knowledge discovery in twitter streaming data. In Discovery Science, pages 1–15. Springer, 2010.

[8] Alec Go, Richa Bhayani, and Lei Huang.Twitter sentiment clas-

sification using distant supervision. CS224N Project Report, Stanford, 2009. URL http://s3.eddieoz.com/docs/sentiment_analysis/Twitter_Sentiment_Classification_using_Distant_Supervision.pdf.

[9] Alec Go, Lei Huang, and Richa Bhayani. Twitter sentiment analysis. CS224N Project Report, Stanford, 2009. URL http://www-nlp.stanford.edu/courses/cs224n/2009/fp/3. pdf.

[10] Barry Kort, Rob Reilly, and Rosalind W Picard. An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In Advanced Learning Technologies, pages 43–436. IEEE Computer Society, 2001.

[11] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and A. Perera. Opinion mining and sentiment analysis on a twitter data stream. Advances in Information and communications technology for Emerging Regions, 13:182–188, Dec 2012.

[12] Brijesh Kumar Baradwaj and Saurabh Pal. Mining educational data to analyze students' performance. International Journal of Advanced Computer Science and Applications, 2(6):63–69, 2011.

[13] Bart Czernicki. Creating data visualizations for analysis. In Silverlight 4 Business Intelligence Software, pages 175–217. Springer, 2010.

[14] Claire Knight. Visualisation effectiveness. In International Conference on Imaging Science, Systems, and Technology, pages 249–254, 2001.

[15] Charles F Harrington, Scott A Gordon, and Timothy J Schibik. Course management system utilization and implications for practice: A national survey of department chairpersons. Technical Report 4, University of West Georgia Richards, 2004. URL http://www.westga.edu/~distance/ojdla/winter74/harrington74.htm.

[16] Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. Analysis of emotion recognition using facial expressions, speech and multimodal information. In Proceedings of the 6th International Conference on Multimodal Interfaces, ICMI '04, pages 205–211. ACM, 2004.

[17] David Callele, Eric Neufeld, and Kevin Schneider. Visualizing emotional requirements. In Fourth International Workshop on Requirements Engineering Visualization, pages 1–10. IEEE, 2009.

[18] Diana Inkpen, Fazel Keshtkar, and Diman Ghazi. Analysis and generation of emotion in texts. Knowledge engineering: principle and technique, 2:3–14, 2009.

[19] David Crystal. English as a global language. Cambridge University Press, 2012.

[20] Dmitry Davidov, Oren Tsur, and Ari Rappoport. Semi-supervised recognition of sarcastic sentences in twitter and amazon. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning, 10, pages 107–116. Association for Computational Linguistics, 2010.

[21] Eric Gilbert and Karrie Karahalios. Widespread worry and the stock market. In Association for the Advancement of Artificial Intelligence, volume 4, pages 59–65, 2010.

[22] E Holly Buttner. How do we "dis" students?: A model of (dis) respectful business instructor behavior. Journal of Management Education, 28(3):319–334, 2004

[23] Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore. Twitter sentiment analysis: The good the bad and the omg! ICWSM, 11:538–541, 2011.

[24] Edward F Gehringer. Applications for supporting collaboration in the classroom. In American Society for Engineering Education. American Society for Engineering Education, 2012.

[25] Fazel Keshtkar and Diana Inkpen. A corpus-based method for extracting paraphrases of emotion terms. In Computational approaches to Analysis and Generation of emotion in Text, pages 35–44. Association for Computational Linguistics, 2010.

[26] Frank Dellaert, Thomas Polzin, and Alex Waibel. Recognizing emotion in speech. In Fourth International Conference on Spoken Language, volume 3, pages 1970–1973. IEEE, 1996.

[27] Gaya Buddhinath and Damien Derry. A simple enhancement to one rule classification. Technical report, University of Melbourne, Australia., Department of Computer Sci-ence & Software Engineering, 2006. URL http://www.buddhinath.net/OtherLinks/ Documents/Improved%20OneR%20Algorithm.pdf.

[28] Gabriela Grosseck and Carmen Holotescu. Can we use twitter for educational ac-tivities. In The fourth international conference on eLearning and software for edu-cation, volume 4, 2008. URL https://adlunap.ro/else/papers/015.-697.1.Grosseck% 20Gabriela-Can%20we%20use.pdf.

[29] Hatice Gunes and Massimo Piccardi. Bi-modal emotion recognition from expressive face and body gestures. Journal of Network and Computer Applications, 30(4): 1334–1345, 2007.

[30] Hugh M Culbertson and Richard D Powers. A study of graph comprehension difficul-ties. Educational Technology Research and Development, 7(3):97–110, 1959.

[31] Haji Binali, Chen Wu, and Vidyasagar Potdar. Computational approaches for emotion detection in text. In Digital Ecosystems and Technologies (DEST), pages 100–107. IEEE, 2010.

[32] Hugh C Davis, Kate Dickens, Manuel Leon Urrutia, Sanchéz Vera, Maria del Mar, and Su White. Moocs for universities and learners an analysis of motivating factors, Apr 2014. URL http://eprints.soton.ac.uk/363714/1/DavisEtAl2014MOOCsCSEDU Final. pdf.

[33] Henning Bar, Erik Tews, and Guido Robling. Improving feedback and classroom interaction using mobile phones. Proceedings of Mobile Learning, pages 55–62, 2005.

[34] Harry L Dangel and Charles Xiaoxue Wang. Student response systems in higher edu-cation: Moving beyond linear teaching and surface learning. Journal of Educational Technology Development and Exchange,, 1(1):93–104, 2008.

[35] Ivon Arroyo, David G. Cooper, Winslow Burleson, Beverly Park Woolf, Kasia Muld-ner, and Robert Christopherson. Emotion sensors go to school. In Artificial Intelligence in Education, volume 200 of Frontiers in Artificial Intelligence and Applications, pages 17–24. IOS Press, 2009.

[36] Jean Aitchison. Language change: progress or decay? Cambridge University Press, 2001.

[37] Jeff Cain, Esther P Black, and Jurgen Rohr. An audience response system strat-egy to improve student motivation, attention, and feedback. American Journal of Pharmaceutical Education, 73(2):1–7, 2009.

[38] Joseph L Fleiss. Measuring nominal scale agreement among many raters. Psychologi-cal bulletin, 76(5):378, 1971.

[39] Jan Geryk. Visual analytics by animations in higher education. In European Confer-ence on e-Learning, pages 565–572. Academic Conferences International Limited, 2013.

[40] Katherine J Denker. Student response systems and facilitating the large lecture basic communication course: Assessing engagement and learning. Communication Teacher, 27(1):50–69, 2013

**Naima Sabir** received his BSCS (2004-2008) from Govt. College University, Faisalabad (GCUF)-Pak. She started her MS (Software engineering) session 2017-2018 (GCUF), with area of Information Mining, Information Management, Information Processing.

**Muhammad Awais** received his Master Degree in Computer Science from the University of Agriculture, Faisalabad, Pakistan, in 2001, MPhil degree in Computer Science in 2004 from the University of Agriculture, MPhil degree in applied information Science in 2008 from the Albert-Ludwigs University, Freiburg, Germany and Ph.D. degree in Robotics from University of Bayreuth, Germany in 2013. In 2005, he joined the Government college university Faisalabad as Lecturer and became Assistant Professor in 2016. His current research interests include Human robot interaction, intention estimation and machine learning.

**Zubair Nabi** received the BS(CS) degree from .C.U.F Faisalabad in Computer Science in 2006 and M.S in 2009. Currently he is doing PhD from GCUF,Faisalabad in Computer Science since 2015.He is faculty member in department of Software Engineering.

**Muhammad Yahya Saeed** received the MSc degree from UAAR, Rawalpindi, in Computer Science in 2003 and M.S degrees in Computer Science from UAF, Faisalabad in 2007. He has also done MSc from UAF, Faisalabad in Statistics in 2000.Currently he is doing PhD from GCUF, Faisalabad in Computer Science since 2015.He is faculty member in department of Software Engineering.