

# Feature Selection for Automatic Answer Extraction from Online Web Forums

Hemant Sharma<sup>\*1</sup>✉, Shyamol Banerjee<sup>2</sup>✉

<sup>1</sup>M.Tech Student, Dept. of CSE, SRCEM

<sup>2</sup>Assistant Professor, Dept. of CSE, SRCEM

[hemantsharma8120@gmail.com](mailto:hemantsharma8120@gmail.com)

[shyamolgwl@gmail.com](mailto:shyamolgwl@gmail.com)

\*Corresponding Author.: [hemantsharma8120@gmail.com](mailto:hemantsharma8120@gmail.com)



This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

This research proposes an efficient method for automatic answer quality classification in online web forums using a Multi-Task BERT-Based deep learning framework. The primary objective is to accurately categorize user responses into low, medium, and high-quality classes by leveraging advanced language representation and relevant content features. Starting with Stack Overflow forum data collection, the methodology moves on to comprehensive preprocessing, exploratory data analysis, and the extraction of syntactic and semantic features. Syntactic features include things like sentence length, punctuation use, code snippets, and TF-IDF vectors and contextual embeddings using spaCy. To reduce redundancy and increase the quality of the model input, we utilised feature selection with RFECV and the principle component analysis (PCA). A two-output-layer multi-task BERT architecture, which the model uses as its basis, can handle major classification and auxiliary tasks simultaneously, allowing it to increase generalisation. After seven training epochs, the trained model achieved remarkable results: 99.54% Training Accuracy, 92.13% Validation accuracy, 92.13% Precision, 92.13% Recall, and 92.13% F1-Score. In addition to providing fast inference, the model typically only takes 0.0071 milliseconds per sample for reactions. In large-scale community-driven QA platforms, these results validate the model's robustness, efficiency, and appropriateness for real-time applications.

**Keywords:** Answer Quality Classification, Multi-Task BERT, Feature Selection, Online Web Forums and Deep Learning

## Introduction

Particularly in technical and community-based fields, online web forums including Stack Overflow, Quora, Reddit, and others have become valuable archives of knowledge in the age of digital information and user-generated content. User-submitted questions and answers abound on these sites create dynamic question-answer (QA) threads that might help automated systems as well as real-time users. But the enormous amount of material, different writing styles, noise, and repetitious comments in the responses make it difficult to find the most pertinent or "correct" response. Here the idea of automatic response extraction—automating the identification and retrieval of the best feasible answer to a given question—becomes essential[1]–[6]. Building such systems depends on feature selection, which is essential for improving interpretability and



performance alike. By means of the identification of the most useful elements influencing response quality, feature selection helps lower dimensionality, decrease noise, increase model accuracy, and lower computing complexity. The process of extracting relevant answers from online forums is known as supervised classification or ranking, and it relies on labels that indicate how relevant or popular an answer is. In addition to the actual text, these forums provide rich metadata such as timestamps, user reputation ratings, up vote count, existence of code snippets, hyperlinks, and answer length. This makes the feature space multi-dimensional and will include as much textual, structural, social attributes. Attention to the proper selection is very important as it can lead to model overfitting and low performance if features are either redundant or irrelevant. So, before building a strong and scalable models we need to know what are the most impactful features. Feature selection can be generally categorized into three commonly accepted approaches: Filter approach, Embedding strategy and Wrapper methods. With respect to the output label, filter techniques such as Mutual Information, Information Gain and the Chi-Square Test are applicable. evaluate the statistical properties of traits to determine their importance. They can ignore the dependencies between features, but they are computationally cheap and agnostic to model. Wrapper approaches assess feature subsets based on the performance of a particular learning algorithm. This category includes methods such as recursive feature elimination (RFE), which are computationally demanding but typically produce superior results. Embedded techniques are methods that use feature selection during the model-training phase. As they learn, algorithms like Lasso Regression and tree-based models (such as Random Forest or XGBoost) automatically rank data by penalising less important details.[7]–[12]. Due to the great performance on correlated and high dimensional data of embedded methods they fit perfectly into the answer extraction framework. For the forum answer extraction, common features are Textual content, linguistic and semantic properties, user-based features and thread-level features. For textual content, we have features like paragraph length, the existence of domain-specific keywords, TF-inverse document frequency (TF-IDF), n-grams, and readability ratings. For example, sentiment polarity, part-of-speech tagging, named entity recognition and the embedding-based representations (e.g., Word2Vec or BERT embeddings) are some semantic features. Things like how many questions a user has answered previously, how many badges they have earned and their reputation all play a part in claiming that you can trust them. At the thread level, things like context of the inquiry and answer matter. Some of this has to do with the amount of information, question formatting and number of competing responses. Different feature categories help the model better understand answer quality in their own unique mechanisms. Deep Learning = Feature Extraction / Selection on steroids Recent developments in Machine learning as well as natural language processing have enabled us to terribly enhance our approaches around feature extraction and selection. Although neural networks can extract hierarchical representations from unlabelled data, there is actually still a need for preliminary feature selection before training as it prevents overfitting—and this becomes increasingly critical in scenarios where labelled data could be problematic. Feature selection can help in boosting interpretability, especially since trust-sensitive applications (e.g. educational systems or medical QA systems) are gaining importance. This is a vital step in constructing accurate automatic answer-extraction systems from online web forums. Transformers benefit from attention mechanisms that automatically weigh certain features more heavily, effectively highlighting input sections relevant to predicting output classes. It aids in inherent dimensionality reduction and hence speeding up learning process via generalisability with noise reduction & interpretability through influencing factor discovery. To get useful results from traditional machine learning models or cutting-edge deep learning architectures in the chaotic, diverse, and user-driven realm of online QA systems, comprehensive and methodical feature selection is required. Smart, context-aware answer retrieval system research is expected to be led by hybrid approaches that integrate human feature engineering with automated deep learning.[13]–[16].

## Literature Review

Quincozes 2023 et.al put forward an ensemble learning-based strategy for selecting and classifying web services. Research into online services is quite young, and ensemble learning has just been popular for



evaluating the aforementioned attributes. The proposed project seeks to assess current best practices in this fascinating area of internet services through research. Utilising the PRISMA framework—which stands for preferred reporting items for systematic reviews and meta-analyses—this research technique surveys the pertinent literature. The results show that ensemble learning has gained more traction in the selected studies published in the past 10 years. Multiple studies demonstrated the frequent examination of classifiers such as naïve bayes (NB), support vector machines (SVM), and others. This ground-breaking research on web service classification lays the groundwork for future studies on web service performance measures. This study found all four of the most popular performance metrics found in the literature: reliability, precision, recollection, and the f-measure.[17].

Meyer 2023 et.al Consistent inspiration and motivation are required during the long and arduous process of changing one's health-related habits. Based on their previous results, conversational agents seem to be able to assist with the transformation. As a therapeutic method, motivational interviewing provides a foundation for change-inducing conversational agents. Unfortunately, motivational interviewing requires a deep grasp of user speech, which is not present in current uses of this therapy method. We fill this data vacuum with the GLoHBCD, a German dataset of natural language data related to health behaviour change. Annotating data obtained from a well-known German weight loss forum, we applied motivational interviewing labels based on theory. Here, we lay out the steps for building a dataset and present the results of our evaluation. The data and the classifiers that come from it could be useful in many different areas of behaviour modification, according to early studies. In order to restore access to the tests and dataset hosted on GitHub, we have developed the necessary code.[18].

Gaikwad 2023 et.al Many individuals enjoy platforms because they provide a space for unrestricted expression. Using social media for recruitment, fundraising, and organisational development, extremist groups primarily aim to spread harmful propaganda and beliefs. In order to rein in the spread of radical ideologies on social media, extremism identification is an absolute must. The current literature on social media extremism identification is limited by competing ideologies, subjective validation methodologies, and binary/tertiary classifications. It is important to evaluate and contrast different datasets, classification algorithms, validation procedures, and tools in order to detect online extremists. We used the PRISMA guidelines for this extensive literature review. Using the Snowballing technique, I gathered 64 articles about studies of extremism. SCOPUS, WoS, ACM, and IEEE were among the databases that provided 31 of these papers. Analytical papers, technical reports, or theses accounted for the remaining thirty-one. Finding extremism on social media is crucial, and the poll shows that these platforms play a role in the growth of online radicalisation. The dearth of publicly available, class-balanced, objective datasets is preventing better identification and classification of social-media extremism, according to the review's conclusion. Our research has shown that there are currently no systems in place to automatically verify the accuracy and completeness of user-defined data sets. The results show that the academic community is biased towards ISIS's ideology at the moment. We found that deep learning-based automated extremism identification systems outperformed the competition. This study lays the groundwork for future research that may yield an automated online tool that the general population can utilise to track down and collect information about radicals. Results of the architectural conceptualisation survey for building an ideology-agnostic dataset of extremism literature with strong data validation techniques for multi classification [19].

Nazah 2022 et.al the challenging operation of determining the nature and traits of the Dark Web cluster of message boards. Due to the unique characteristics of Dark Web data, current approaches relying on continuously tagged contents are both costly and inefficient. Thus, it is necessary to have a strategy in place in case labelled forums and relevant information are not always easily accessible. So, we provide an unsupervised model that combines clustering and decision tree methods to detect and describe Dark Web forums. By presenting the features in an understandable form, the proposed method helps law enforcement authorities and cyber threat intelligence systems investigate any suspected data breach or illegal activity on



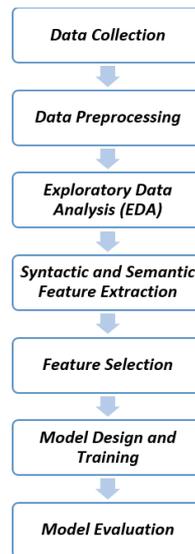
the Dark Web forums. We conducted thorough research before putting our model to the test using actual data collected from DarkWeb communities. Our suggested model accurately describes Dark Web communities, with an F1 score of 98%. Based on the findings of the experiments, the suggested methodology could help the cyber threat intelligence and law enforcement community build a smart database that can detect data breaches and illegal activities on the Dark Web.[20].

Humaira 2022 et. al The AISAS context is very important for showing viewers while buying in this digital age. A key component of AISAS is the recommendation that consumers do two things: collect information and share it. Using postpositivist theory in conjunction with descriptive research designs, case study research methodologies, and current data or digital artefacts obtained from Sociolla's archives is the research paradigm for this work. The 2020 Avoskin cosmetics catalogue, as well as the Sociolla social networking sites saluran Instagram, Soco by Sociolla (Soco.id), and Sociolla's media website Sociolla (sociolla.com) carried the news. With the help of the Sociolla online forum, the results of this study make it possible for AISAS vendors and customers to share and receive relevant information. Features of Sociolla include e-commerce, involvement with online discussion forums, digital auditors, kosmetik with greater knowledge, and review writing.[21].

**Table 1. Literature Summary**

Author/year	Methodology	Research gap	Findings
Jain 2021 [22]	Review-based analytical technique looking at OSN risks, remedies, and difficulties.	Not enough defence mechanisms against changing OSN hazards and user weaknesses.	OSNs are rapidly used for data; so, more privacy protections are very necessary.
Tseng 2021 [23]	mixed-methods investigation of forums exposing techniques of intimate relationship monitoring.	Insufficient IPS research focussing on offenders limits appropriate preventive measures.	Attackers publish IPS techniques openly; forums expose trends in monitoring tools.
Khan 2020 [24]	Lexical feature SVM-based classification for forum answer the extraction process.	Lexical feature SVM-based classification for forum answer the extraction process.	In online forum conversations, Linear SVC accurately labels pertinent responses.
Shrestha 2020 [25]	Analysing language and network aspects, unsupervised RNN or anomaly detection	Few studies on using network and language elements together to identify depression.	Combined network and language traits successfully identify online sad people.
Martin 2020 [26]	Early suicide thought detection on Reddit using LSTM-CNN model	Few studies on advanced deep learning early suicidal ideation detection.	LSTM-CNN model finds suicidal publishes on Reddit site rather successfully.

## Methodology



**Fig 1 Proposed Flowchart**

**A. Data Collection:** This study uses KaggleHub, more especially the "60k Stack Overflow Questions with Quality Rate" dataset by Imoore. This dataset comprises Stack Overflow questions together with a binary label denoting their quality. Kagglehub.dataset\_download is used for retrieval guarantees the most recent version of the dataset is obtained. Using Pandas, load the training and validation files (train.csv and valid.csv). Six thousand questions for validation and eighteen thousand questions for training assist keep computers running smoothly and make sure that evaluations are fair by using only a small part of the data. These questions have a lot of important linguistic parts, like the question title, body, and metadata like timestamps. This dataset is the starting point for looking into automated answer extraction using machine learning, thus it's important to check the data's reliability, representativeness, and labeling quality. In training classification models, the labels (Y) tell you if a query is good or bad. This setup is like a real-life situation where varied quality questions come up in forums, and the goal is to predict and pick the best ones to answer.

**B. Data Preprocessing:** The first step in preprocessing is to look at the structure of the dataset and use.info() to summarize the types of data. It also needs to change the Creation Date into a datetime format and get the Year of publishing and other useful information. Libraries like Seaborn and Matplotlib are used to look at class distribution and posting trends over time. The most crucial step is to turn unprocessed text into inputs that may be used in a model. We use Hugging Face Transformers' Bert Tokenizer from the collection to do this. BERT tokenization makes attention masks, splits the text into subword tokens, and adds special tokens such [CLS] and [SEP]. This makes sure that the assumptions of the BERT model are still valid and helps to deal with the fact that forum messages can be different lengths. The title and body of each post are probably tokenized and encoded, and then reduced or padded to fit BERT's token limit. This preparation makes sure that there is as little unnecessary noise as possible, such as extended HTML or code fragments, which keeps the semantic richness. This step makes sure that the model's inputs are all the same and lets you get temporal and category metadata to help you choose features.

**C. Exploratory Data Analysis (EDA):** The exploratory data analysis (EDA) stage is when you look at the dataset in depth to find any unusual patterns or trends. The study can discover imbalances by looking at the class distribution of the quality labels (Y) initially. Making histograms for Title\_length and

Body\_length helps you see how content sizes are spread out, which lets you see if questions that are too lengthy or too short affect quality classification. Bar graphs can let you see how often tags are used with inquiries, which will show you the most popular technical topics on the forum. These new facts let us decide on feature engineering, like content length or tag count, which could be very good signs of postings that deserve an answer. Temporal analysis based on the CreationDate lets you see trends in posting over time that may be linked to changes in the quality of questions or the popularity of tags. EDA's goal is to give you a fundamental understanding of the data's structure and composition so that you may better preprocess and extract features. This step makes sure that the dataset is good overall and gives early hints about which features might be important for predicting high-quality responses to forum posts.

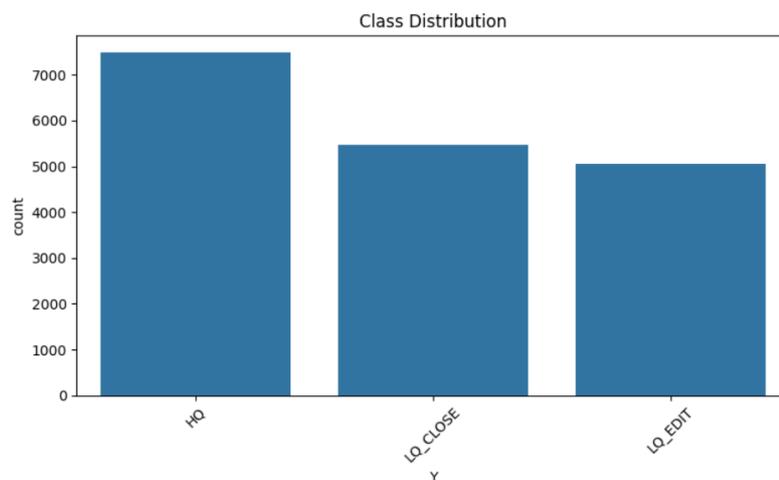


Fig 1: Class Distribution

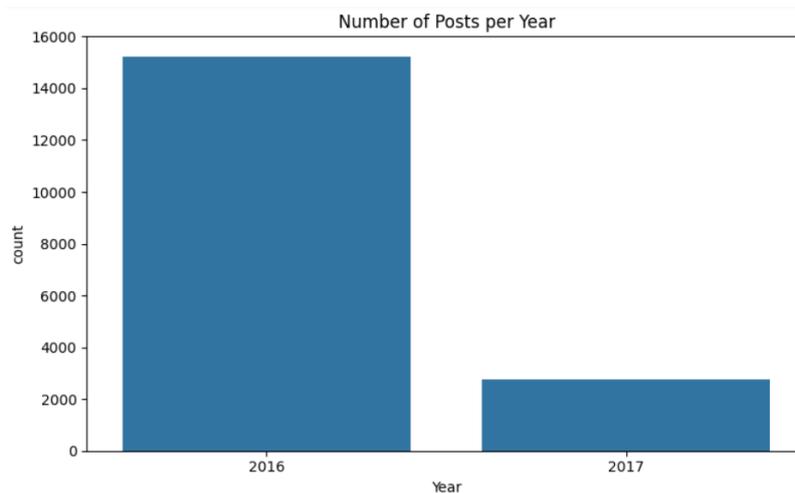


Fig 2: Number of Posts per year

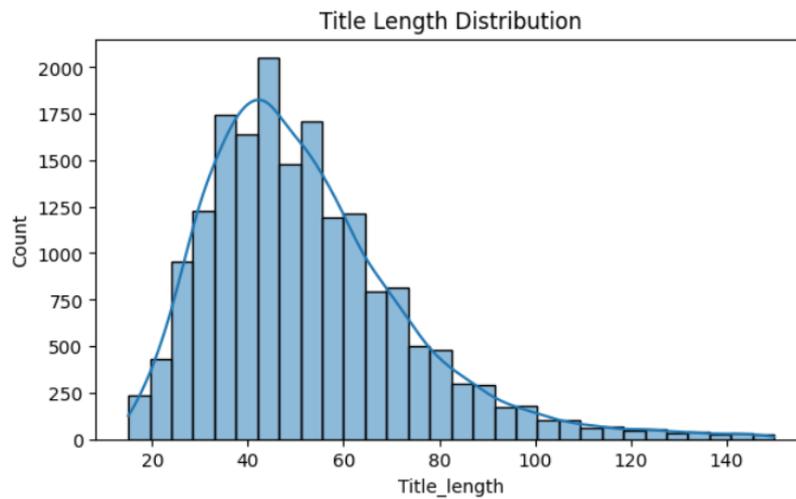


Fig 3: Title length Distribution

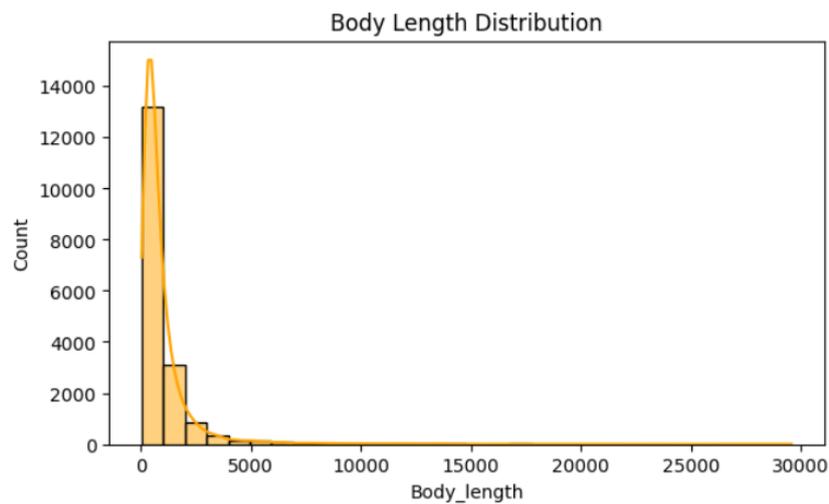


Fig 4: Body length Distribution

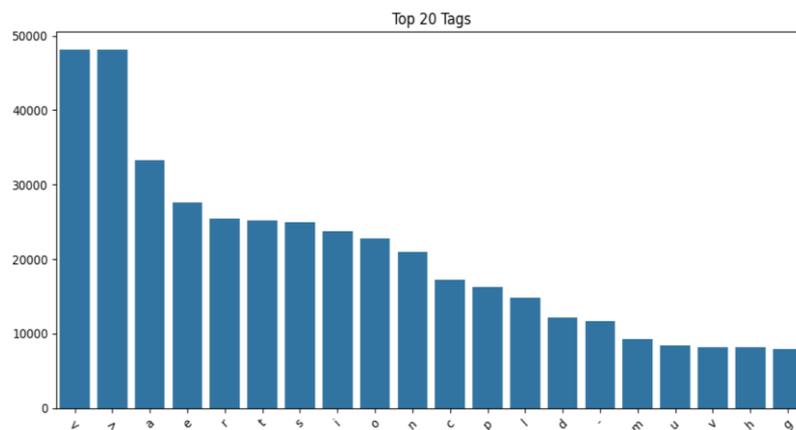


Fig 5: Top 20 tags

**D. Syntactic and Semantic Feature Extraction:** In this work, feature extraction happens in two ways: syntactic and semantic. Syntactic components are basic structural metrics like post length and tag count



that are calculated from the title and content of each post combined. These are simple yet strong signs of how rich and focused a post is. Semantic feature extraction is more complicated and involves turning text into high-dimensional numbers. We employed spaCy embeddings and TF-IDF vectorizing as our two strategies. The TF-IDF method highlights vocabulary that is distinctive to a certain field by comparing the relevance of individual terms in each post to the importance of all the terms in the corpus. Using spaCy's `en_core_web_sm` model lets you make rich word embeddings that capture contextual semantics at the same time. These embeddings are averaged per document to generate a fixed-size representation. Syntactic features, TF-IDF vectors, and SpaCy embeddings were then put together to make a whole feature matrix that shows both the surface structure and the underlying meaning of each post. This hybrid representation aims to combine deep contextual understanding with easy-to-understand information, which is important for accurately figuring out which forum questions are most likely to get good answers.

**E. Feature Selection:** Dimensionality reduction and feature selection are necessary because combined syntactic, TF-IDF, and semantic embeddings create a lot of dimensions. Principal Component Analysis (PCA) was first used to reduce noise and redundancy while keeping the feature set variation. This phase of compression cuts the feature space down to 100 parts. Then, Recursive Feature Elimination with Cross-Validation (RFECV) is used with a Random Forest Classifier to pick the most useful features. Using F1-score (weighted) as the evaluation metric over 5 stratified folds, RFECV looks at feature subsets by removing the least important ones one at a time. This method strikes a mix between model generalisability and feature relevance. The method finally generates an optimal feature subset with low overfitting risk and good predictive power while lowering computation time. Using downstream model training, the selected features are applied, which improves the classifier's performance and cost-effectiveness. This study relies heavily on feature selection since it narrows down the input space to focus on the most useful syntactic and semantic signals, which makes it easier to classify high-quality forum discussions more accurately.

**F. Model Design and Training:** The fundamental model for this work is a multi-task BERT-based architecture created in PyTorch. A pretrained bert-base-uncased model is used to encode text. The model has two output heads: an auxiliary head that may be used for other binary tasks like engagement or response rate, and a core classifier that estimates the quality of the question. Dropout regularization and pooling BERT outputs as inputs to the linear classifier layers are part of the design of the architecture. Cross-Entropy Loss is used for the primary work, and Binary Cross-Entropy Loss is used for the extra work. The optimizer utilized is Adam with a learning rate of  $2e-5$ . Data loaders that work with tokenized input sequences that are all the same length for batch processing do the training. Each training cycle includes forward and backward passes, gradient updates, and calculations of accuracy in the middle. The multi-task setup encourages generalization by requiring shared feature representations that are useful for related goals. This layout not only makes it easier to classify things correctly, but it also helps people understand the meaning of forum messages better, which is important for finding good information in complicated online debates.

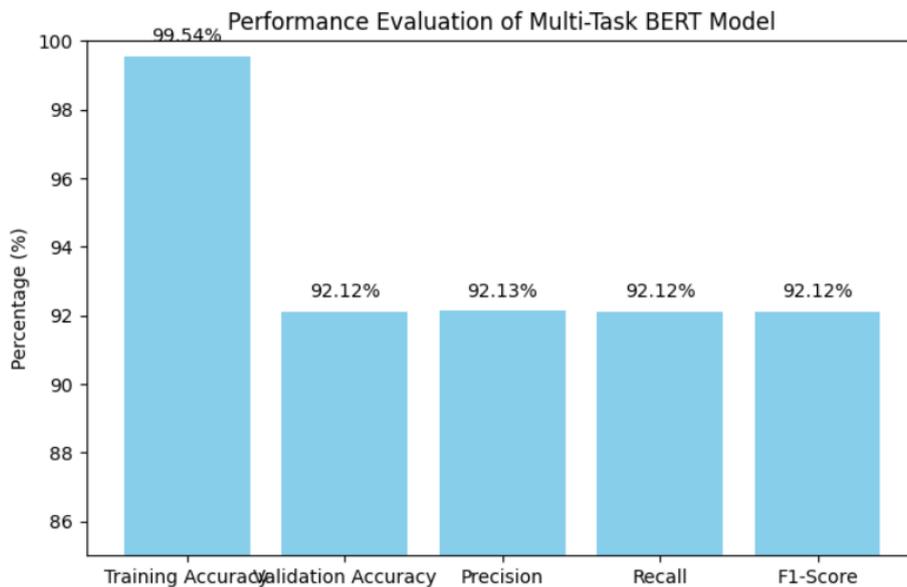
**G. Model Evaluation:** Use the standard classification metrics—accuracy, precision, recall, and F1-score—to test the model on the validation set. Overall, these metrics illustrate how well the model works on courses that aren't balanced. The evaluation function uses a weighted average to compare the forecasts made by the trained BERT model with the ground truth labels. This takes into consideration the imbalance of the labels. The categorization report's per-class statistics tell you if the model is better at finding high-quality or low-quality posts. The average response time per sample of the model is also calculated. This shows how efficient it is, which is very important for real-time use. In this scenario, F1-score is more important than accuracy since it balances false positives and false negatives. Adding extra tasks helps to validate findings more consistently and makes learning more regular, even with fake labels.

This rigorous evaluation approach shows that the model not only learns from training data but also works well on new posts. This makes the answer-extraction method more robust in dynamic web forum settings.

## Result & Discussion

**Table 2. Performance Evaluation of the Multi-Task Bert-Based Answer Quality Classification Model**

Metric	Value
Final Training Accuracy	99.54%
Final Validation Accuracy	92.12%
Precision	92.13%
Recall	92.12%
F1-Score	92.12%
Average Response Time/Sample	0.0071 sec



**Fig 7: Performance Evaluation of Multi-Task BERT Model**

The Multi-Task BERT-Based Answer Quality Classification Model did a great job of sorting answers posted on internet forums into different quality groups when it was time to do so. As training came to a close, the model showed remarkable generalisability, with validation accuracy at 92.10% and training accuracy at 99.54%. There was a 92.13% accuracy rate, a 92.12% recall rate, and an F1-score of 92.13% for the model. Using these efficient and comprehensive metrics, we can see that the model accurately differentiates between low-, medium-, and high-quality responses. Surprisingly, it improved on all classes over time and kept a perfect score on class 2 responses (high-quality comments) for every epoch. With an average response rate per sample of just 0.0071 seconds, the model is obviously perfect for application in situations requiring immediate results. Based on these results, it is reasonable to assume that online discussion platform quality classification is best handled by a multi-task BERT structure that uses syntactic and semantic data.



## Conclusion

This study demonstrates the efficacy of a Multi-Task BERT-Based deep learning model in automatically categorising the quality of responses in online web forums. Through the integration of syntactic and semantic feature extraction with robust feature selection methods like PCA and RFECV, the model is able to extract meaning from forum comments beyond the surface level. While the performance of the model and its generalisation ability on different quality classes are significantly improved by adopting a multi-task structure with two learning objectives. The final training accuracy at the end of 30 epochs was 99.54%, while the validation accuracy was at 92.12%, and thus this model produced high evaluation metrics through extensive training, and testing over many epochs. This conclusion was further supported by strong metrics for precision (92.13%), recall (92.12%) and the F1-score (92.12%). It can produce correct results in as little as 7.1 milliseconds per frame, making it accurate enough for real-time application and fast enough to be used properly in practice. This pattern demonstrates the power of transformer-based architectures and advanced processing techniques on complex tasks that require high levels of user involvement to get right. Overall, the study presents both a pragmatic and computable means of automating solution assessment which can aid in improving user experience & content moderation on large online discussion forums. User interaction metrics and multi-platform creation are two areas for future development.

## Acknowledgment

Write your acknowledgment here using the same text format.

## Authors' Declaration

- Conflicts of Interest: None.

-

## References

- [1] S. Jaki, T. De Smedt, M. Gwó, R. Panchal, A. Rossa, and G. De Pauw, "Online Hatred of Women in the Incels . me Forum : Linguistic Analysis and Automatic Detection State of the Art," 2019.
- [2] D. P. R. De Lima, M. A. Gerosa, T. U. Conte, and J. F. D. M. Netto, "What to expect , and how to improve online discussion forums : the instructors ' perspective," vol. 0, 2019.
- [3] S. Razniewski, J. Z. Pan, and G. Weikum, "Commonsense Properties from Query Logs and Question Answering Forums," 2018.
- [4] T. Zhang, G. Upadhyaya, A. Reinhardt, H. Rajan, and M. Kim, "Are Code Examples on an Online Q & A Forum Reliable? A Study of API Misuse on Stack Overflow," pp. 886–896, 2018, doi: 10.1145/3180155.3180260.
- [5] T. K. F. Chiu and T. K. F. Hew, "Factors influencing peer learning and performance in MOOC asynchronous online discussion forum," vol. 34, no. 4, pp. 16–28, 2018.
- [6] Z. Mohamadi, "Studies in Educational Evaluation Comparative effect of online summative and formative assessment on EFL student writing ability," vol. 59, no. July 2017, pp. 29–40, 2018.
- [7] S. Ji, C. P. Yu, S. Fung, S. Pan, and G. Long, "Supervised Learning for Suicidal Ideation Detection in Online User Content," vol. 2018, 2018, doi: 10.1155/2018/6157249.
- [8] T. Internet, B. Chen, Y. Chang, F. Ouyang, and W. Zhou, "CO," no. October 2019, 2018, doi: 10.1016/j.iheduc.2017.12.002.
- [9] S. Zhang, X. I. N. Zhang, H. U. I. Wang, L. Guo, and S. Liu, "Multi-Scale Attentive Interaction Networks for Chinese Medical Question Answer Selection," vol. 6, pp. 74061–74071, 2018.
- [10] Q. M. Zhong and H. Norton, "Educational Affordances of an Asynchronous Online Discussion Forum



for Language Learners,” vol. 22, no. 3, pp. 1–19, 2018.

- [11] E. F. Bloomfield and D. Tillery, “The Circulation of Climate Change Denial Online : Rhetorical and Networking Strategies on Facebook,” 2018.
- [12] H. Shing, S. Nair, A. Zirikly, M. Friedenberg, H. Daum, and P. Resnik, “Expert , Crowdsourced , and Machine Assessment of Suicide Risk via Online Postings,” pp. 25–36, 2018.
- [13] T. Chakrabarty, C. Hidey, S. Muresan, K. Mckeown, and A. Hwang, “AMPERSAND : Argument Mining for PERSuasive oNline Discussions,” 2017.
- [14] L. Zheng, H. Wang, and S. Gao, “Sentimental feature selection for sentiment analysis of Chinese online reviews Sentimental feature selection for sentiment analysis of Chinese online reviews,” no. March 2015, 2017, doi: 10.1007/s13042-015-0347-4.
- [15] D. Hazarika, E. Cambria, and R. Zimmermann, “CASCADE: Contextual Sarcasm Detection in Online Discussion Forums,” 2016.
- [16] J. Glass and B. Randeree, “SemEval-2015 Task 3: Answer Selection in Community Question Answering,” 2015.
- [17] M. Hasnain, I. Ghani, S. R. Jeong, and A. Ali, “Ensemble learning models for classification and selection of web services: A review,” *Comput. Syst. Sci. Eng.*, vol. 40, no. 1, pp. 327–339, 2022, doi: 10.32604/CSSE.2022.018300.
- [18] S. Meyer and D. Elswiler, “GLoHBCD: A Naturalistic German Dataset for Language of Health Behaviour Change on Online Support Forums,” *2022 Lang. Resour. Eval. Conf. Lr. 2022*, no. June, pp. 2226–2235, 2022.
- [19] M. Gaikwad, S. Ahirrao, S. Phansalkar, and K. Kotecha, “Online Extremism Detection: A Systematic Literature Review with Emphasis on Datasets, Classification Techniques, Validation Methods, and Tools,” *IEEE Access*, vol. 9, no. C, pp. 48364–48404, 2021, doi: 10.1109/ACCESS.2021.3068313.
- [20] S. Nazah, S. Huda, J. H. Abawajy, and M. M. Hassan, “An Unsupervised Model for Identifying and Characterizing Dark Web Forums,” *IEEE Access*, vol. 9, pp. 112871–112892, 2021, doi: 10.1109/ACCESS.2021.3103319.
- [21] N. Humaira, “ISAS (Attention, Interest, Search, Action, Share) Model of Cosmetics Marketing Communication on Online Beauty Forum (Case-Study: Avoskin Marketing on Sociolla),” *MEDIALOG J. Ilmu Komun.*, vol. 4, no. 1, pp. 186–200, 2021, doi: 10.35326/medialog.v4i1.1031.
- [22] A. K. Jain, S. R. Sahoo, and J. Kaubiyal, “Online social networks security and privacy: comprehensive review and analysis,” *Complex Intell. Syst.*, vol. 7, no. 5, pp. 2157–2177, 2021, doi: 10.1007/s40747-021-00409-7.
- [23] E. Tseng *et al.*, “The tools and tactics used in intimate partner surveillance: An analysis of online infidelity forums,” *Proc. 29th USENIX Secur. Symp.*, pp. 1893–1909, 2020.
- [24] A. Khan *et al.*, “Machine Learning Approach for Answer Detection in Discussion Forums: An Application of Big Data Analytics,” *Sci. Program.*, vol. 2020, 2020, doi: 10.1155/2020/4621196.
- [25] A. Shrestha, E. Serra, and F. Spezzano, “Multi-modal social and psycho-linguistic embedding via recurrent neural networks to identify depressed users in online forums,” *Netw. Model. Anal. Heal. Informatics Bioinforma.*, vol. 9, no. 1, 2020, doi: 10.1007/s13721-020-0226-0.
- [26] F. Martin, C. Wang, and A. Sadaf, “Facilitation matters: Instructor perception of helpfulness of facilitation strategies in online courses,” *Online Learn. J.*, vol. 24, no. 1, pp. 28–49, 2020, doi: 10.24059/olj.v24i1.1980.