Comparative Analysis of Machine Learning and Deep Learning Models for Aspect-Based Sentiment Analysis in Education

ABSTRACT

Aspect-Based Sentiment Analysis (ABSA) has emerged as a powerful technique for analyzing studentfeedbackineducationalsettings,providingadeeperunderstandingofsentimentslinkedtospecificaspectssuchascoursecontent,instructorperformance,assessmentqualityandtechnology support.Unliketraditional sentiment analysis, ABSA enables granular insights by extracting multiple aspects from a singlereview and assigning sentiments to each aspect independently. This study evaluates the performance oftraditional Machine Learning (ML) models, including Logistic Regression (LR), Support Vector Machines(SVM), Naïve Bayes (NB), Random Forest (RF) and Gradient Boosting (GB), alongside advanced DeepLearning (DL) models such as Multi-Layer Perceptron (MLP), Recurrent Neural Networks (RNN), LongShort-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT). Thefocusisonaddressingthechallengeofhandlingmultipleaspectsperreviewandperformingaspect-specificsentimentclassification.ExperimentalresultsdemonstratethatBERTsignificantlyoutperformsothermodelsin both tasks, offering superior precision,recalland F1-scores. Notably, BERT excels in handling complex, multi-aspect feedback, providing more accurate sentiment classification for each aspect.These findingshighlighttheimportanceofleveragingadvancedmodelstoanalyzeeducationalfeedbackeffectively,enablinginstitutionstoimplementtargetedimprovementsinkeyareasoflearningand teaching.

Keywords:AspectExtraction,BERT,StudentFeedback,LongShort-TermMemory,RandomForest,Sentiment Classification.

1.INTRODUCTION

The rapid expansion of online education has resulted in an overwhelming amount of student feedback,whichservesasanessentialsourceofinformationforimprovingeducationalservices.Asmoreinstitutionsmovetowarddigitallearningenvironments,platformssuchasMassiveOpenOnlineCourses(MOOCs)andother e-learning initiatives are receiving substantial feedback from students across the globe. TraditionalSentimentAnalysis(SA)methodstypicallyclassifyfeedbackaspositive,negativeorneutral.However, thismethod falls short of identifying and analyzing specific aspects of the learning experience, such as coursecontent, instructor performance and technology support. To address this, Aspect-Based Sentiment Analysis(ABSA)hasemergedasamorenuancedapproachthatlinkssentimenttospecificaspectsofthefeedback(Hussainetal.,2024).

Unlike traditional SA, ABSA allows for a more granular analysis by categorizing feedback into differentaspects, providing actionable insights for administrators and educators. For instance, feedback aboutcourse content can be classified separately from feedback on instructor performance or technology issues(Ngwira et al., 2023). This granular analysis is especially beneficial for institutions looking to improvespecificelementsoftheircoursesorteachingstrategies.

The growing complexity and volume of data in educational settings make ABSA a crucial tool for educationaldataanalytics.WhileNaturalLanguageProcessing(NLP)andArtificialIntelligence(AI)havebeeninstrumentalinimprovingsentimentclassificationandaspectextraction,theshifttoonlinelearningandtheincreasingvarietyofstudentfeedbacknecessitatemoreadvancedtools.Thishasledtotheadoptionof bothMachineLearning(ML)andDeepLearning(DL)modelstoprocesslarge,unstructuredfeedbackdata.These models not only improve the accuracy of aspect extraction but also allow for deeper sentimentanalysis,providing more detailedinsightsintothestudentlearning experience(Ngocetal.,2021).

While several studies have already employed ABSA using advanced NLP and DL techniques, these studiesoften fail to account for the multiple aspects present in each review. Instead, they typically assign a singleaspect category to the entire feedback, along with a single sentiment classification for that aspect (Ngwiraet al., 2023). However, student reviews frequently contain multiple aspects (e.g., course content, teachingquality, technology support) with distinct sentiments tied to each one. By not distinguishing between theseaspects,previousmodelslimitthedepthoftheiranalysisandtheabilitytoaddressspecificconcernsraisedby students.

This gap in existing research underscores the importance of the current study, which aims to extract multipleaspects per review and assign sentiment to each identified aspect separately. This approach allows for amuch finer-grained analysis of educational feedback, which is essential for improving specific componentsoftheeducationalexperience,suchascoursematerials,instructoreffectivenessandtheuseoftechnologyin course delivery. This study not only aims to compare the performance of traditional ML models and DLalgorithms for ABSA but also to explore how deep learning models, particularly BERT, can be used tohandlemultipleaspectsperreviewandperformaspect-specificsentimentclassification.

The primary objective of this study is to compare the performance of several ML and DL algorithms in twocritical tasks of ABSA: aspect extraction and sentiment classification. This study evaluates a range of MLmodels, including Logistic Regression (LR), Support Vector Machines (SVM), Naïve Bayes (NB), GradientBoosting (GB) and Random Forest (RF) and compares them to advanced DL models such as Multi-LayerPerceptron (MLP), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and BidirectionalEncoder Representations from Transformers (BERT). By evaluating these algorithms, this paper aims todeterminewhichmodelsarebest suitedtohandlethecomplexityofeducational feedbackdata.

Therestofthepaperisstructuredasfollows:Section2presentstheliteraturereview,summarizingpreviousresearch on ABSA, including the use of ML and DL models in education and highlighting gaps in handlingmultiple aspects per review. Section 3 details the methodology, explaining the models, algorithms andpreprocessing techniques used in this study. Section 4 discusses the experiments and results, followed bySection 5, which provides a discussion of the findings and Section 6 concludes the paper with keytakeawaysandfutureresearchdirections.

2| LITERATUREREVIEW

ABSA has become an increasingly valuable tool for analyzing student feedback in educational contexts,offering a more detailed understanding of specific aspects of the learning experience. Recent research inthe field highlights how ABSA has been applied to educational datasets to extract meaningful insightsrelated to course content, instructor performance and the use of technology in learning environments. Forexample, Kastrati et al. (2020) demonstrated the application of a weakly supervised ABSA framework toanalyze student feedback from MOOCs. Their work showed how ABSA could be applied to large datasetswith minimal reliance on manually annotated data, significantly reducing the need for expensive human-labeleddatasets.

Moreover, Heryadi et al. (2022) applied ABSA to student feedback on online learning programs, revealingimportant insights that could help educational institutions improve their course offerings. Their studyhighlighted the ability of ABSA to extract specific feedback related to various aspects of online courses,including content quality, teaching methods and course structure. Similarly, Alassaf and Qamar (2020) usedABSAforanalyzingArabic-languageeducationaltweets,whichunderscoredtheimportanceofmultilingualaspectextractionindiverseeducationalcontexts.

ML models have been integral to ABSA, especially for tasks such as sentiment classification and aspectextraction. Traditional ML algorithms such as LR and SVM are effective in analyzing smaller datasets andhandlingsimplerbinaryormulti-classsentimentclassificationtasks.However,thesemodelsfacechallengeswhenappliedtolarge-scale,complexdatasetswheredeeperlinguisticstructuresneedtobe understood.SVM,forinstance,hasbeenshowntoperformwellonsmallerdatasetsbutoftenunderperforms in more complex scenarios (Sindhu et al., 2019). However, ML models still face limitationsin terms of their inability to handle the deep contextual dependencies that exist in educational feedbackdata. For instance, models like NB and LR fail to capture the subtle relationships between words in longer,more complexsentences,thuslimiting theireffectivenessinABSAtasks(Kastratiet al.,2020).

Ontheotherhand,ensemblemodelslikeRFandGBarehighlyeffectiveinreducingoverfittingandincreasingtheaccuracyofsentimentclassification.Thesemodelsworkbycombiningthepredictionsofmultipleclassifierstoimproveoverallperformance.RF,inparticular,hasshownstrongresultsinaspectextractiontasks, where nuanced relationships betweenwords need tobe captured fromfeedbackdata(Edalati etal.,2022).

Deep learning models, particularly RNN and LSTM networks, have shown significant improvements inABSA by addressing the limitations of traditional machine learning methods. RNNs and LSTMs excel atcapturing sequential dependencies in text data, which is crucial for sentiment classification and aspectextraction in feedback, where sentiments may be spread across multiple sentences or contexts. Thesemodelshavebeenproveneffectiveinhandlingthesequentialnatureoffeedback,especiallyinreviewsthatcontainrich,context-dependentsentiments(Sindhuetal.,2019).

However, BERT, a transformer-based model, has surpassed traditional ML and DL methods due to itsability to process bidirectional context. This ability allows BERT to understand the relationships betweenwords in both directions, making it especially effective for ABSA in educational settings. Alshaikh et al.(2023) found that BERT significantly outperformed other models in both aspect extraction and sentimentclassificationtasks,particularlywhenanalyzingfeedbackfromMOOCsandtraditionaleducationalsettings

. Additionally, BERT's capability to capture long-range dependencies and its ability to process complexfeedback from students in diverse educational contexts have made it the model of choice for ABSA ineducation.

In summary, while BERT has proven effective in ABSA tasks, particularly for extracting sentiments fromstudentfeedback,mostexistingstudieshavefocusedonassigningasingleaspectcategoryandsentimentclassificationperreview.Thisstudyaimstobridgethisgapbyextractingmultipleaspectsfromeachreviewand performing aspect-specific sentiment classification, offering a more detailed and accurate analysis ofeducationalfeedback.

3. METHODOLOGY

3.1 Dataset

This study utilized a synthetic dataset consisting of 1,500 student feedback records, with 100 recordscorresponding to each of the 15 (2^4 - 1) possible aspect categories. Each record contains six columns:*Review*, *Aspect_Category*, *Course*, *Instructor*, *Technology* and *Assessment*. The *Review* column containssingle or multiple sentences, each of which may address one or more aspects. For example, a singlesentence might express sentiments about both the course content and the instructor's performance, whileanother sentence in the same review might address technology and assessment. The *Aspect_Category*fieldrepresentscombinationsoffourbroadaspects:course,instructor,technologyandassessment.Thereare 15 possible *Aspect_Category*values, ranging from individual aspects (e.g., *course*) to combinations ofmultiple aspects (e.g., *course, instructor, technology, assessment*). The dataset includes columns for*Course*, *Instructor*, *Technology* and *Assessment*, each representing the corresponding sentiment polarity(positive,negativeorneutral)fortheaspectspresentinthe*Review*.Thisdatasetprovidedacomprehensive foundation for both aspect extraction and sentiment classification tasks. Each record captured detailedfeedback, enablinganuanced analysisofthevariouselementsofthelearningexperience.

3.2 DataPreprocessing

Before training the ML and DL models, several preprocessing steps were applied to clean and normalizethe textdata.The followingstepsoutlinethe preprocessingprocess:

- **Lowercasing:** All text was converted to lowercase to ensure that the models treat words like“Course”and“course”asidentical,eliminatingcasesensitivityissues.
- **Removal of Numbers:** Numerical values, such as dates and scores, were removed using regularexpressions. Numbers were not considered relevant for sentiment or aspect extraction and werethusexcludedfromtheanalysis.
- **Removal of Punctuation:** Punctuation marks were removed, except for full stops and commas,asthesecanbeimportantforsentenceboundariesandsentimentclassification.Specialcharactersandunnecessarysymbolswerealsostrippedawaytoavoidnoiseinthedata.

After these text cleaning transformations were applied, the cleaned text was vectorized using the TermFrequency-Inverse Document Frequency (TF-IDF)technique.The TF-IDFvectorizertransformed thetextualdataintonumericalfeaturevectors,whichcouldthenbeusedfortrainingbothmachinelearninganddeep learning models. TF-IDF ensured that frequently occurring words in a review were given appropriateweight relativetotheirfrequency acrosstheentiredataset.

3.3 DataSplitting

Thedatasetwassplitinto80%fortrainingand20%fortestingtoensurethemodelshadasufficientamountofdataforlearning,whileleavingaportionofthedataforunbiasedevaluation.Thissplitiscommonlyusedin MLandDLstudiestoprovide arobustestimateofmodelperformance.

- Training Set(80%): Used for fittingtheMLand DL models.
- TestingSet (20%):Usedfor evaluating theperformance ofthe trained models.

Thesplitwasstratifiedtoensurethatthedistributionofaspectcategoriesandsentimentclassesremainedconsistent across both the training and testing datasets. This approach prevents any imbalance that couldskewthemodel’sperformanceonunseendata.

3.4 AspectExtraction

The aspect extraction task was treated as a multi-label classification problem, where each review could beassociated with multiple aspects. The goal was to predict one or more aspects relevant to each review.To handle overlapping aspects in a review (i.e., when multiple aspects are addressed in a single review), each aspect was treated as an independent label. A multi-output classifier approach was used, where each aspect category (course, instructor, technology, assessment) was predicted separately for each review.
Toachievethis,**LogisticRegression(LR)**wasusedasthebasemodel,implementedwithinaMultiOutputClassifiertohandlemultipleaspectcategoriessimultaneously.LRwaschosenforitssimplicity,interpretability and effectiveness in multi-class and multi-label classification tasks. As a probabilistic linearmodel,LRestimatesthelikelihoodofatargetoutcomebymodelingtherelationshipbetweeninputfeaturesand a binary or categorical response variable. Introduced by Cox (1958), LR has been a foundationalalgorithm in machine learning, particularly for tasks requiring probabilistic predictions. In this study, eachaspect (e.g., course, instructor, technology, assessment) was treated as an independent label and themodel was trained on the TF-IDF features of the reviews to predict the presence of one or more aspects.The model was trained using 80% of the data (training set) and its performance was evaluated on theremaining 20% (test set). Metrics such as precision, recall and F1-score were calculated for each aspectcategorytoassessthemodel’seffectivenessinpredicting relevantaspects.

In addition to Logistic Regression, various ML and DL models were employed for aspect extraction, eachbringingdistinctstrengths:

- **Support Vector Machines (SVM)**: A model known for its ability to handle high-dimensional dataefficiently.SVM is particularly effective in text classification tasks, leveraging hyperplanes toseparateclassesinfeaturespace(CortesandVapnik,1995).
- **Naïve Bayes (NB)**: A probabilistic classifier based on Bayes' theorem. While computationallyefficient and easy to implement, NB assumes feature independence, limiting its performance incapturingcomplexdependenciesintextdata(Lewis,1998).
- **Random Forest (RF) and Gradient Boosting (GB)**: Ensemble learning methods that combinepredictions from multiple classifiers to enhance robustness and accuracy. RF constructs multipledecision trees during training, while GB iteratively improves weak learners to minimize classificationerrors. These models are particularly suitable for multi-label classification tasks where multipleaspectsmayco-occurinasinglereview (Breiman,2001;Friedman,2001).
- **Multi-Layer Perceptron (MLP)**: A feedforward neural network capable of capturing non-linearpatterns in text. MLP is widely used in text classification tasks due to its adaptability and ability tomodelcomplexrelationshipsbetweenfeatures(Rosenblatt,1958).
- **Recurrent Neural Networks (RNN)**: Designed to handle sequential data, RNNs are particularlyusefulincapturingdependencieswithintext.Theyprocessinputsequentially,maintaininginformation frompriorwordsto betterunderstandcontext(Rumelhartetal.,1986).
- **Long Short-Term Memory (LSTM)**: A variant of RNN capable of learning long-term dependencies.LSTMs are especially suited for understanding the relationships between multiple aspects in longerreviews,astheymitigatethe vanishinggradient problem (HochreiterandSchmidhuber, 1997).
- **BidirectionalEncoderRepresentationsfromTransformers(BERT)**:Astate-of-the-arttransformer-basedmodelthatprocessestextbidirectionally,capturingnuancedrelationshipsbetweenwords.Pre-trainedonextensivecorpora,BERThasdemonstratedexceptionalperformance inextractingaspectsfromunstructured studentfeedback(Devlinetal.,2019).

3.5 SentimentClassification

Afterextractingtherelevantaspectsfromeachreview,thenextstepwastoclassifythesentiment(positive,negative or neutral) associated with each identified aspect. Sentiment classification was performed for each aspect independently. This allowed the handling of conflicting sentiments(e.g., praise for course content and criticism of the instructor) by assigning separate sentiment labels (positive, negative, or neutral) to each aspect.Separate classifiers were trained for eachaspect category to predict the sentiment for reviews where that aspect was present. For instance, a LRclassifier was trained for the "course" aspect on reviews where the "course" aspect was identified. The TF-IDF features were used as input to predict whether the sentiment was positive, negative or neutral for thataspect. Similar classifiers were trained for other aspects like "instructor," "technology," and "assessment."Thisallowedeachclassifiertofocusonly on the relevantreviewsforeachaspect.

Inaddition to LR, thefollowingMLand DL modelswere usedforsentiment classification:

- **SVM:**Leveragingitsstrengthinhigh-dimensionalfeaturespaces,SVMeffectivelyclassifiedsentimentforeachaspect,particularlyinsparseTF-IDFrepresentations.
- **NB:**Whileefficientandcomputationallylightweight,NBwaslesseffectiveforsentimentclassificationduetoitsinabilityto model complexdependenciesbetweenwords.
- **RF and GB:** Both ensemble methods improved sentiment classification by combining the outputsof multipleclassifiers,enhancing robustnessandreducingoverfitting.
- **MLP:** TheneuralnetworkarchitectureofMLPcapturedcomplexpatternsintext,improvingsentimentclassification performanceforaspectswithsubtlecontextualvariations.
- **RNN and LSTM:** These modelsexcelledinunderstandingsequentialdependenciesintext,enabling

themto capturecontext-dependent sentimentseffectively.

- **BERT:** With its pre-trained bidirectional transformer architecture, BERT achieved state-of-the-artresults in sentiment classification. Its ability to process both word-level and sentence-level contextallowedittoidentify nuancedsentimentstiedtospecificaspects.

3.6 EvaluationMetrics

The models were evaluated on their performance in two key tasks: aspect extraction and sentimentclassification. Foraspect extraction, precision, recalland F1-scorewere calculatedfor each aspect categorytomeasurehoweffectivelythemodelsidentifiedtherelevantaspectswithinstudentreviews.Forsentimen tclassification, detailed classification reports were produced for each aspect, including precision, recall andF1-scores for the sentiment classes (positive, negative and neutral). The evaluation results provided acomprehensiveoverview of the strengths and weaknesses of each model in handling both aspect extractionand sentiment classification, enabling a clear comparison of their capabilities in aspect-based sentimentanalysis.

4. RESULTS

4.1 AspectExtraction

The results for aspect extraction underscore the strengths and limitations of both traditional ML and DLmodels in identifying multiple aspects within a single review. Among all models evaluated, BERT emergedasthetopperformer,achievingperfectprecision,recallandF1-scoresof1.00.Thisoutstandingperformance demonstrates BERT's ability to effectively model complex relationships within unstructuredtext, particularly in reviews where multiple aspects coexist, such as feedback on course content, instructorperformance and technology support. By leveraging its bidirectional attention mechanism, BERT was abletoidentifyrelevantaspectswithunmatched accuracy,avoidingfalsepositivesandnegatives.

DL models such as LSTM and MLP closely followed BERT, achieving F1-scores of 0.99. These modelsdemonstratedtheirstrengthinsequentialdataprocessing,capturingdependenciesbetweenwordsthatar eessential for understanding the nuances of multi-aspect feedback. However, their reliance on sequentialprocessing meant they could not fully match the efficiency and accuracy of BERT, particularly in handlingreviewswithoverlappingorsubtleaspectmentions.

Table1.ResultsofAspectExtraction

Model	Precision	Recall	F1-Score
LR	0.98	0.98	0.98
SVM	0.99	1.00	0.99
NB	0.87	0.97	0.92
RF	0.99	0.99	0.99
GB	0.99	0.98	0.98
MLP	0.99	1.00	0.99
RNN	0.98	0.97	0.97
LSTM	0.99	0.99	0.99
BERT	**1.00**	**1.00**	**1.00**

Traditional ML models such as RF and SVM also performed strongly, achieving F1-scores of 0.99. Thesemodels excelled in aspect extraction due to their robust feature-handling capabilities, particularly whenpairedwitheffectivefeatureengineeringtechniques.GBandLRachievedslightlylowerF1-scoresof0.98,indicating that while they are effective for structured data, they are less adaptable to the complexities ofmulti-aspect text. NB lagged behind, with an F1-score of 0.92, highlighting its limited ability to modelinterdependenciesinfeedbackcontainingmultipleaspects.

The results of aspect extraction are summarized in Table 1, which presents the precision, recall and F1-scores for each model. Figure 1 provides a visual comparison of the models' performances, highlightingthe superiorcapabilitiesofBERTandthecompetitiveresultsofRF,SVMandLSTM.

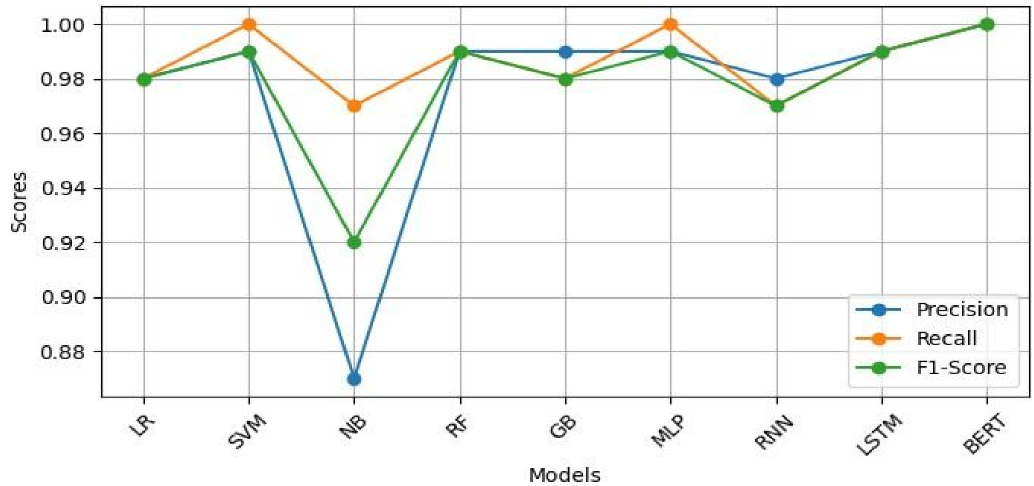

Fig.1.ModelscomparisonforAspect Extraction

4.2 SentimentClassification

The sentiment classification task introduced additional challenges, as the models needed to associatesentiments (positive, negative or neutral) with each identified aspect. Once again, BERT demonstrated itssuperiority, achieving an F1-score of 0.97. Its ability to handle aspect-specific sentiment classification wasparticularly notable, as it excelled at capturing contextual dependencies for each aspect, even in reviewswith mixed sentiments across aspects. This performance further highlights BERT's ability to address thecomplexity of educational feedback, where sentiments are often distributed across multiple sentences oraspects.

Table2.Resultsof SentimentClassification

Model	Precision	Recall	F1-Score
LR	0.92	0.91	0.89
SVM	0.94	0.94	0.93
NB	0.80	0.88	0.84
RF	0.94	0.94	0.94
GB	0.95	0.95	0.95
MLP	0.94	0.94	0.93
RNN	0.90	0.90	0.90
LSTM	0.94	0.95	0.94
BERT	**0.97**	**0.97**	**0.97**

Other DL models, such as GB, LSTM and RF, performed strongly, achieving F1-scores of 0.95, 0.94 and0.94, respectively. These models showed robust capabilities in contextual sentiment analysis, particularlyforwell-structuredfeedback.SVMandMLPalsodemonstratedcompetitiveperformance,withF1-scoresof
0.93 and 0.94. However, their reliance on fixed feature representations limited their effectiveness incapturing nuanced and context-dependent sentiments. LR (F1-score of 0.89) and NB (F1-score of 0.84)struggledsignificantlyinthistask,reflectingtheirinadequacyinhandlingthecontextualintricaciesrequiredforaspect-specificsentimentclassification.

The detailed metrics for sentiment classification are presented in Table 2, which includes precision, recalland F1-scores for all models. Fig. 2 visually compares the performance of the models in sentimentclassification,clearlyshowingthesuperiorperformanceofBERTalongsidethestrongresultsofGB,LSTMand RF.

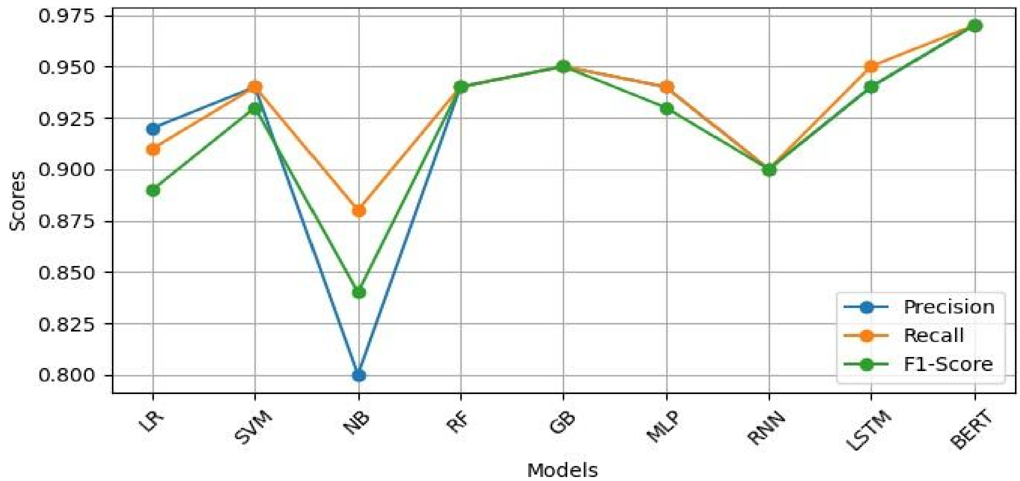

Fig.2.ModelscomparisonforSentimentClassification

5. DISCUSSION

The results of this study reveal several important insights into the challenges and opportunities of ABSA ineducational feedback. The ability to extract multiple aspects from a single review and assign sentiments toeach aspect independently is a critical requirement for meaningful analysis and the performance of themodelsinthisstudyhighlightsthe relativestrengthsandweaknessesofdifferentapproaches.

BERT’soutstandingperformanceacrossbothtasksunderscoresitscapacitytoaddresstheuniquechallenges posed by multi-aspect feedback. Its bidirectional context modeling enables it to capture subtlerelationships between words, making it particularly effective in handling feedback with overlapping orconflicting sentiments. For instance, a single review might praise course content while criticizing instructorperformance. BERT’s architecture allows it to identify and classify sentiments for each aspect separately,offeringalevelofgranularity thatisessentialforactionableinsights.

DL models such as LSTM and MLP also demonstrated their utility, particularly in aspect extraction, wheretheir ability to model sequential dependencies resulted in strong performance. However, their reliance onsequential processing meant that they were less efficient than BERT in handling complex feedback whereaspectsandsentimentsaredistributedacrosssentences.

Traditional ML models, particularly RF and SVM, performed well in aspect extraction, largely due to theirability to handle structured data and capture relationships through feature engineering. However, theirlimitationsbecameapparentinsentimentclassification,wheretheabilitytomodelcontextualdependenciesis crucial. NB, the weakest performer across both tasks, highlights the inadequacy of simplistic models inaddressingthenuancednatureofABSAtasks.

The findings also highlight the importance of using task-specific classifiers for sentiment classification. Bytraining separate classifiers for each aspect, the study ensured that models could focus on the uniquecharacteristics and requirements of each aspect. This approach proved particularly effective in improvingthe accuracy of sentiment classification, as it allowed for a tailored analysis of aspects such as coursecontent,teachingquality andtechnologysupport.

Overall,theresultsemphasizetheneedforadvancedmodelslikeBERTtohandlethecomplexitiesofmulti-aspecteducational feedback.Byprovidinggranularinsightsintostudentsentiments,thesemodelsenable educational institutions to identify specific areas for improvement and make data-driven decisions toenhancethelearningexperience.

6. CONCLUSION

This study provides a comprehensive analysis of ML and DL algorithms for ABSA, with a specific focus onextractingmultipleaspectsperreviewandperformingaspect-specificsentimentclassification.ThefindingsdemonstratethatBERTconsistentlyoutperformsothermodelsinbothtasks,achievingthehighestprecision,recallandF1-scores.Itsabilitytocapturebidirectionalcontextandmodelcomplexdependenciesallows it to effectively handle the challenges of multi-aspect feedback, making it the most reliable andeffectivemodel forABSAineducationalsettings.

DL models such as LSTM and MLP showed strong performance, particularly in aspect extraction, but theirrelianceonsequentialprocessingmadethemlessefficientthanBERTinhandlingthecomplexitiesofmulti-aspectreviews.TraditionalMLmodelslikeRFandSVMwerecompetitiveinaspectextractionbutstruggledwithsentimentclassification,highlightingtheirlimitationsinmodelingcontextualdependencies.NBconsistentlyunderperformed,reflecting itsinabilitytoaddressthe intricacies ofunstructured feedback.

Thestudyunderscoresthetransformativepotentialof ABSAineducationalinstitutions.Bylinkingsentiments to specific aspects within feedback, ABSA enables targeted improvements in areas such ascourse content, teaching quality, assessment strategies and technology support. The ability to analyzemulti-aspectfeedbackprovidesgranularinsightsthattraditionalsentimentanalysismethodscannotachieve.

Future research could extend these findings by exploring the application of BERT and other transformer-basedmodelstomultilingualandcross-domaindatasets.Additionally,hybridapproachesthatcombinethestrengths of traditional ML and DL techniques could further enhance the efficiency and scalability of ABSAsystems. By continuing to refine and expand these methodologies, educational institutions can unlock thefullpotentialoffeedbackanalysis,drivingimpactfulanddata-drivenimprovementsinthelearningexperience.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.

2.

3.

REFERENCES

1. Alassaf, M., and Qamar, A. M. (2020). Aspect-based sentiment analysis of Arabic tweets in theeducation sector using a hybrid feature selection method. In 2020 14th International Conferenceon InnovationsinInformationTechnology(IIT)(pp.178-185).IEEE.

2. Alshaikh, K. A., Almatrafi, O., and Abushark, Y. B. (2023). BERT-based model for aspect-basedsentimentanalysisforanalyzingArabicopen-endedsurveyresponses:Acasestudy.IEEEAccess.

3. Breiman,L. (2001).Randomforests.MachineLearning,45,5-32.

4. Cox, D. R. (1958). The regression analysis of binary sequences. Journal of the Royal StatisticalSocietySeriesB:StatisticalMethodology,20(2),215-232.

5. Devlin,J.,Chang,M.W.,Lee,K.,andToutanova,K.(2019).BERT:Pre-trainingofdeepbidirectional transformersforlanguageunderstanding.NAACL.

6. Edalati,M.,Imran,A.S.,Kastrati,Z.,andDaudpota,S.M.(2022).Thepotentialofmachinelearningalgorithms for sentiment classification of students' feedback on MOOC. In Intelligent Systems andApplications: Proceedings of the 2021 Intelligent Systems Conference (IntelliSys) Volume 3 (pp.11-22).SpringerInternationalPublishing.

7. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals ofStatistics, 1189-1232.

8. Heryadi, Y., Wijanarko, B. D., Murad, D. F., Tho, C., and Hashimoto, K. (2022). Aspect-basedsentimentanalysisforimprovingonlinelearningprogrambasedonstudentfeedback.In2022IEEEInternational Conference on Cybernetics and Computational Intelligence (CyberneticsCom) (pp.505-509).IEEE.

9. Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8),1735–1780.

10. Hussain, T., Yu, L., Asim, M., Ahmed, A., and Wani, M. A. (2024). Enhancing e-learning adaptabilitywithautomatedlearningstyleidentificationandsentimentanalysis:Ahybriddeeplearningapproachforsmarteducation.Information,15(5),277.

11. Kastrati, Z., Imran, A. S., and Kurti, A. (2020). Weakly supervised framework for aspect-basedsentimentanalysisonstudents' reviewsofMOOCs.IEEEAccess, 8,106799-106810.

12. Lewis, D. D. (1998). Naive (bayes) at forty: The independence assumption in information retrieval.In EuropeanConferenceonMachineLearning (pp.4-15).Berlin,Heidelberg:SpringerBerlinHeidelberg.

13. Ngoc, T. V., Thi, M. N., and Thi, H. N. (2021). Sentiment analysis of students' reviews on onlinecourses: A transfer learning method. In Proceedings of the International Conference on IndustrialEngineeringandOperationsManagement(pp.306-314).

14. Ngwira,B.,Gobin-Rahimbux,B.,andSahib,N.G.(2023).Adeep-learningframeworkforanalysingstudents'reviewinhigher education. ComputationalIntelligenceandNeuroscience, 2023(1),8462575.

15. Rosenblatt,F.(1958).Theperceptron:Aprobabilisticmodelforinformationstorageandorganizationinthebrain.Psychological Review,65(6),386.

16. Sindhu, I., Daudpota, S. M., Badar, K., Bakhtyar, M., Baber, J., and Nurunnabi, M. (2019). Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation. IEEEAccess, 7,108729-108741.