U.Porto Journal of Engineering, 8:2 (2022) 95-107
ISSN 2183-6493
DOI: 10.24840/2183-6493_008.002_0008

Received: 2 November, 2021
Revised: 15 December, 2021
Published: 28 April, 2022

Automatic Grading of Student's Presentation Skills based on PowerPoint Presentation and Audio

J. G. Borade[1], A. W. Kiwelekar[2], L. D. Netak[3]
[1]Research Scholar, Department of Computer Science and Engineering, Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra 402103, India (jyoti.borade81@gmail.com) ORCID 0000-0002-6062-7249; [2]Faculty of Computer Science and Engineering, Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra 402103, India (awk@dbatu.ac.in) ORCID 0000-0002-3407-0221; [3]Faculty of Computer Science and Engineering, Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra 402103 (ldnetak@dbatu.ac.in) ORCID 0000-0001-7388-9822

Abstract
Creating and delivering a presentation for Project, Seminar, and Course work is an important academic activity included in the curricula of undergraduate engineering studies. The presentation should be graded based on the presentation skills, accuracy, and authenticity of the contents covered in the presentation. Educational institutes use rubrics to assess the presentation skills on different grounds, which is a cumbersome task for the teacher when the strength of students is significant. Our main objective is to automatically grade the students' presentation skills in terms of PowerPoint presentations and the student's confidence. The proposed system describes a method and dataset designed to automate grading students' presentation skills. Our research study is divided into two parts. In the first part, the PowerPoint presentation features corresponding to text appearance, tables, charts, images, footer, and hyperlinks are extracted to grade PowerPoint presentations. At the same time, Mel-frequency Cepstral Coefficients, Mel Spectrogram, and Chroma features are extracted from the students' audio to identify confidence in the second part of the study. The audio is recorded at the presentation time. Feature extraction programs are implemented in python using Python-pptx and Librosa library. The tree-based feature selection method is used to remove the irrelevant features. Random Forest Ensemble model gives 100 % accuracy while predicting the grade of PowerPoint presentations. Multilayer Perceptron model gives 88% accuracy while predicting the confidence level of the students. The output of both models is combined to grade the students' presentation skills. The quadratic Weighted Kappa (QWK) score is 0.82, which indicates a significant similarity between automated and human-rated scores.

Author Keywords. PowerPoint Slides, Audio Files, Machine Learning Models, Python-pptx, Librosa.

Type: Research Article

1. Introduction

Automated evaluation is an emerging technology. Evaluating students' knowledge plays a crucial role in effective teaching and learning. An automated grading system may reduce the paper load of teachers and assessment-related issues. Teachers spend extensive time assessing students' technical and non-technical performance in essay types of works like answer sheets, articles, research papers, programming code, thesis, reports, and PowerPoint presentations (Borade, Kiwelekar, and Netak 2022; Borade and Netak 2021). For the same work, different evaluators offer different scores. Hence, grades are not assigned to the students based on their learning abilities. It may happen due to various reasons like:

- A human rater may be holistic and harsh; he may be moderate and liberal; he may be unskilled and not follow the evaluations' instructions.
- Grading such essay work takes a long time; it is stressful and laborious.
- Manual scoring is susceptible to inconsistencies and inaccuracies from human raters.

Grading students' work is an essential task in any course because it is a kind of feedback for the students about learning the concepts. It measures the performance and learning achievements of the students.

There is no research on rating the quality of a student's presentation skills. Some researchers may think it is a minor part, but the students need to impart good presentation and technical skills. Delivering an effective presentation is a soft skill that students can use in their careers. It is essential to achieve soft skills in the student's state because it enables students to adjust to the frustrations and challenges that they will face in the future. Project, Seminar, Course presentation are all included in the curricula to develop the presentation and technical skills of the students. Our aim is only about automated grading of the presentation skills. Automated grading of presentation skills is an original project, which is very useful to evaluate the quality of students' PowerPoint presentations and the student's confidence. The teacher needs only to grade the presentation based on the contents or topic expertise of the students. Also, our system will undoubtedly assist students in identifying shortcomings in their presentation skills. They can test their presentation skills and improve till satisfaction using our system before the actual presentation.

2. Related Work

Several automated grading systems are developed to evaluate different types of work. In 1968, Page and Paulus used statistical techniques to correlate writing style with grade and ignored actual text. Some researchers have used Natural Language Processing (NLP) to extract various linguistic features from text and applied various machine learning (ML) techniques to grade it. Deep neural networks have automatic feature extraction capability. Hence, it has replaced feature engineering based on NLP. Automated grading systems are developed for automatically reviewing research articles, grading programming assignments, poems, short answers, and long answers. Leng, Yu, and Xiong (2019) developed an automated system to review research papers. They used Hierarchical Recurrent Convolutional Neural Network (HRCNN), Multilayer Perceptron (MLP), and Convolutional Neural Network (CNN) models to get semantic, grammatical, and innovative features respectively from the research papers. These extracted features are merged into one vector and fed to MLP, the predictive model, to predict the final review score. The predictive model is trained and tested on the research articles from openreview. Parihar et al. (2017) presented a system for evaluating programming assignments in an introductory programming course. Automatic evaluation is done by using the number of test cases passed, the correctness of the program, the time taken to solve programming assignments, and the number of successful compilations.

Many researchers have designed evaluation systems for various languages. Al-Jouie and Azmi (2017), Azmi, Al-Jouie, and Hussain (2019) and Bashir et al. (2018) used Latent Semantic Analysis and Deep Learning techniques for natural language understanding in the Arabic language. Ajitiono and Widyani (2016) developed automated grading of explanatory answers in the Indonesian language. They used NLP to extract language-based features. Walia, Josan, and Singh (2019) developed grading of answers in the Punjabi language using Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM). Agung Putri Ratna et al. (2018) developed a grading system for the Japanese language examinations. They used the

Winnowing algorithm to check the similarities between the texts. It is based on hashing technique. Peng et al. (2010) developed a system to evaluate Chinese essays using the Latent Semantic approach. Janda et al. (2019) used syntactic, semantic, and sentiment analysis to evaluate essays. They extracted various syntactical features of the essays. Sentences are represented as nodes of the graph, and semantic similarities are represented as a weight of the edges. They derived some features from these graphs. All these systems have used lexical features, syntax, grammar-related features, and semantic features of the languages. NLP is used to extract a set of language features like number of words, number of paragraphs, POS, usage of a preposition, conjunction, adjective, verb, adverb, noun, number of sentences, vocabulary, and spelling checking (Fazal, Hussain, and Dillon 2013). Linguistic features can also be extracted automatically by applying deep learning approaches. George, Sijimol, and Varghese (2019) and Surya, Gayakwad, and Nallakaruppan (2019) extracted features using deep models. Voigtlaender, Doetsch, and Ney (2016) used deep models to recognize characters and applied ML techniques to perform grading tasks.

Dar and Khaki (2013) have developed an SVM classifier model to recognize the speaker's emotion using Mel-frequency cepstral coefficients (MFCC). MFCC represents the speaker's vocal tract information. Deshmukh et al. (2019) presented an emotion recognition system in the Hindi and Marathi languages. Harb and Chen (2003) and Yucesoy and Nabiyev (2013) have used MFCC features to identify the gender of the speakers. Zeng et al. (2006) have developed a model to recognize the male and female voices. Ladde and Deshmukh (2015) have developed multiple classifier systems for emotion recognition and gender identification from the audios. CNN is used to classify gender and age from audio signals (Dat and The Anh 2019; Kuchebo et al. 2021). Kattel et al. (2019) presented chroma features extraction using different methods. Automated systems are developed for emotion recognition and gender identification from audio signals. However, no work is found to check the speaker's confidence from its audio. Confidence is an essential parameter of presentation skills.

3. System Architecture

We have completed our research study in two parts. The first is to ensure that PowerPoint slides are of good quality. The second is to recognize the students' confidence from the audio recorded during the presentation. The first part of our system has evaluated PowerPoint presentations depending on various presentation features used by the student. The second system has recognized the student's confidence depending on various audio features.

We have gathered 150 PowerPoint presentations designed by third and fourth-year engineering undergraduate students as a part of their Seminar and Project activity. Additionally, during the delivery of a presentation, each student's audio is captured with their permission. We have implemented a feature extraction program to extract features from PowerPoint presentations and audios using Python-pptx and Librosa library. Machine learning algorithms require a labeled dataset. A panel of 3 teachers has independently graded PowerPoint presentations and confidence. The majority of the grades are assigned in the dataset as output labels with the permission of the experts. We have prepared two datasets; a dataset of PowerPoint presentations and the students' audio. Experts have used their experience and expertise to grade the students' presentation skills. Figure 1 and Figure 2 depict our system architecture for automatically grading students' presentation skills.

3.1. Evaluating the grade of PowerPoint presentations

As shown in Figure 1, solid lines show training, and dotted lines show the testing phase of the model. Labeled PowerPoint presentation feed as input to our system. Each PowerPoint presentation is represented by an input feature vector of size 23, corresponding to 23 features listed in Table 1 in section 4.1. These features attract the audience during the delivery of the presentation and grades PowerPoint presentations into Excellent, Very Good, Good, and Fair categories. Removal of irrelevant features helps avoid over-fitting with a small number of data samples. Hence, we applied ExtraTree Classifier and Linear Discriminant Analysis techniques to remove irrelevant features and reduce feature dimensions. Machine learning models are used to establish a relationship between input features and output grades assigned by the evaluators. Machine learning models are tested by feeding Ungraded PowerPoint presentations to them.

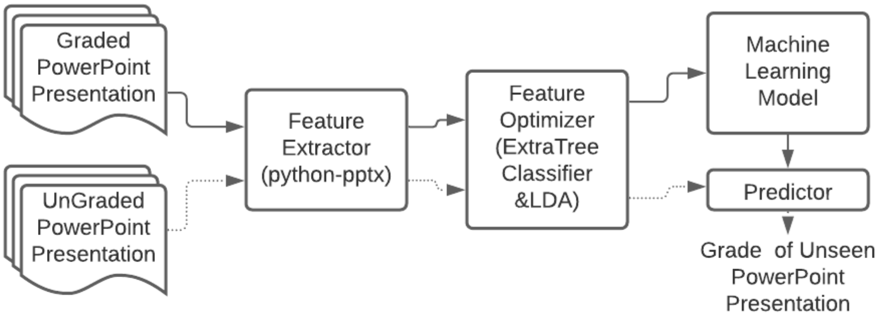

Figure 1: Grading of the PowerPoint presentation

3.2. Evaluating the student's confidences

The sound of the confident student is loud and clear. It indirectly reflects good presentation and language skills. Such students can attract the audience and establish good interaction. The student's confidence is an important parameter, needs to be considered in the presentation skills. Hence, the evaluation of the confidence of the student is essential.

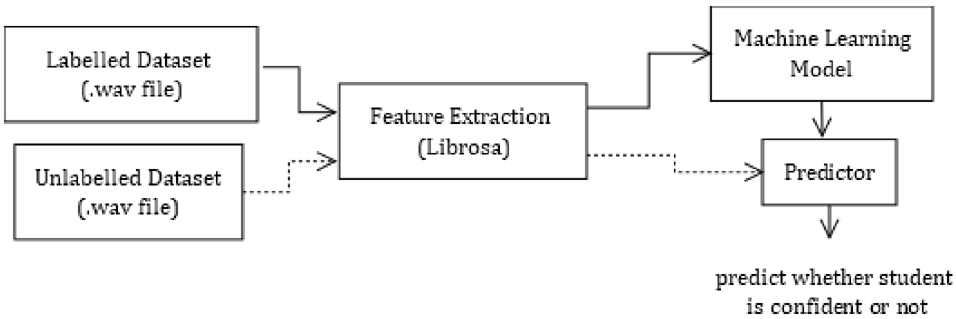

Figure 2: Evaluating the confidence of the student from the audio file

Figure 2 shows our system's architecture to identify students' confidence levels. The solid lines show training, and the dotted lines show the testing phase of the model. Labeled Audio feed as input to our system. Audio features like MFCC, Chroma, and Mel Spectrogram categorize speakers as confident or non-confident. Linear Discriminant Analysis techniques remove irrelevant features and reduce feature dimensions. Machine learning models are used to establish a relationship between input audio features and output grades assigned by the evaluators. The machine learning model is tested by feeding unseen audio data to them.

4. Feature Extraction

In this section, we have presented various features extracted from the PowerPoint presentations and audio files.

4.1. Feature extraction from the PowerPoint presentations

We have collected PowerPoint presentations prepared by the students. They prepared PowerPoint slides using presentation software like MS Office, Python-pptx library. We have extracted features that check the efforts put by students to prepare PowerPoint presentations and not for the topic and communication skills. The PowerPoint presentation can be made effective by using the features related to the appearance of text, graphics, footer, and hyperlink (Table 1).

Feature No.	Feature Name	Description
1	Image	Checks the presence of an image, diagram, or figure.
2	Chart	Checks presence of a chart.
3	Table	Checks presence of a table.
4	Textbox	Checks whether text box is used or not.
5	Body	Checks whether the body is present in the slides.
6	Placeholder	Checks the presence of placeholder in the slides.
7	Title	Checks use of the title in the slides.
8	Centre title	Checks use of Centre title in the slides.
9	Subtitle	Checks use of Subtitles in the slides.
10	Footer	Indicates presence of footer.
11	Date	Checks use of date.
12	Slide number	Indicates the use of slide numbering in the slides.
13	Number of slides	Represents the number of slides present in the presentation.
14	Number of hyperlinks	Indicates the number of hyperlinks used in the presentation.
15	Hyperlink	Checks presence of hyperlinks.
16	Maximum font size	Indicates the maximum text font size used in the paragraph.
17	Minimum font size	Indicates the minimum size of text font used in the paragraph.
18	Bold	Checks presence of bold formatting.
19	Underline	Indicates the presence of underline formatting.
20	Italic	Indicates the use of italic text.
21	Number of font types	Indicates the number of font types used for the text.
22	Number of font sizes	Indicates the number of font sizes used for the text.
23	Number of font colors	Indicates the number of font colors used for the text.

Table 1: List of the PowerPoint presentation features extracted using Python-pptx library

These features capture the information about the appearance of the texts in terms of colors, font size, and font types. They attract the audience during the delivery of the presentation. The use of graphics conveys information more effectively. Hyperlink provides smooth navigation across presentations. We can add date, topic name, institute name, candidate name, and the slide number using Header and Footer. Hence, we have extracted these features to evaluate the quality of the PowerPoint presentations. Program is implemented for features extraction from the PowerPoint presentations using the python-pptx library.

4.2. Features extraction from the audio files

The audio signal is analyzed using Librosa, a python library. Program is implemented for features extraction from the audio files using Librosa. We have extracted the following features from the .wav audio files to check whether a student is confident or not. Graphical representation of these features depicts the difference between the confident and non-confident speakers (Figure 3 and Figure 4).

1. **Mel-Frequency Cepstral Coefficients (MFCC):** The human's vocal tract determines the sound. Any sound produced can be precisely described if the vocal tract shape is identified appropriately. The boundary of the time power spectrum of the audio signal represents the vocal tract, and MFCC appropriately depicts it. Boundaries of the time power

spectrum are sufficient to express the differences between phonemes. MFCC is used to recognize these phonemes. Figure 3 depicts the difference between MFCCs of confident and non-confident speakers. There are 39 features of MFCCs.

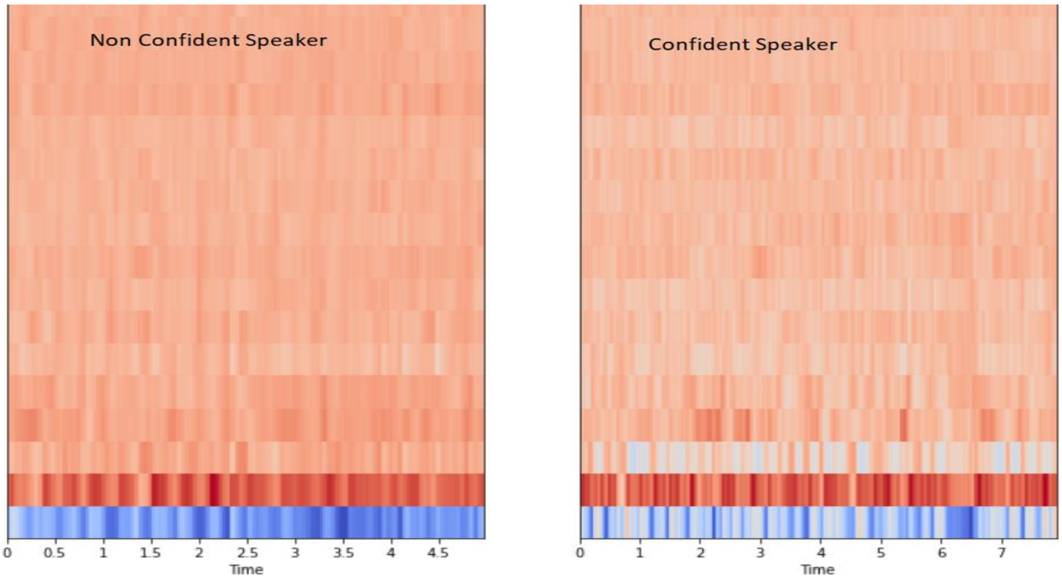

Figure 3: MFCC of a non-confident and confident speaker

2. **Chroma Features:** The twelve different pitch classes represent chroma features. Pitch is the property that allows the categorization of sounds as higher or lower. Pitch represents harmonic and melodic features from the audio. The pitch of the audio signal is found by the frequency with which the audio wave trembles. Chroma features have 12 elements. Each element indicates the energy of each pitch class.
3. **Mel Spectrogram Frequency:** It visualizes the strength of the audio. It allows us to see the shape and form of the audio by visualizing audio and the pressure that these sound waves cause. A confident speaker's voice is loud, and it resonates in an upward direction compared to a non-confident speaker's voice (Figure 4). A confident and non-confident voice takes a different shape and can predict whether the speaker is confident or non-confident. A logarithmic transformation of a signal's frequency is the Mel scale frequency. Mel Spectrograms are sound spectrograms that display sound on the Mel scale rather than the frequency domain. Humans can detect minor variations in pitch at low frequencies. In this scale, features become closer to the audibility of humans hear. Mel Spectrograms is popular in machine learning since it simulates human perception.

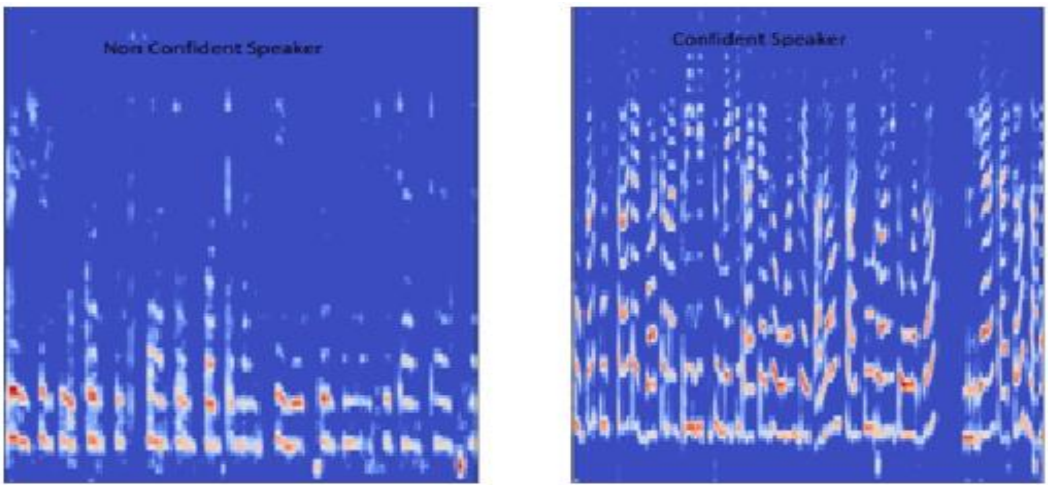

Figure 4: Mel Spectrogram of a non-confident and confident speaker

5. Model Developments

We have used various supervised machine learning techniques to build machine learning models. Machine learning models establish a relationship between input features and output. The number of samples is the same in both datasets, but the number of input features is different. Hence, the hyperparameters used for the models in the first part of the study will not work for the second part. We have briefly described the models with hyperparameters used in our work.

1. **Decision Tree (DT):** Decision nodes and leaf nodes are the two types of nodes in the DT. Leaf nodes represent the expected class value for a label: Excellent, Very Good, Good, and Fair in the first part and Confident and Non-Confident in the second. Decision nodes examine each feature for specified criteria, i.e., 'gini' and 'entropy', to split a node into classes. We have used 'entropy' in our experimentation.
2. **Multi-Layer Perceptron (MLP):** Optimization algorithms are used in neural networks for weight optimization during training. We have used the regularization parameter 'alpha' to control the overfitting of the model. In the first part of our study, we have used the Limited-Memory Broyden-Fletcher-Goldfarb-Shanno, i.e., 'lbfgs' optimizer, and set regularization parameter 'alpha' to 1e-5. We have used 18 hidden layers, with 8 neurons and a 'tanh' activation function. In the second part of our study, we have used an adaptive moment estimation, i.e., 'adam' optimizer, and set regularization parameter 'alpha' to 0.05. We have used 350 hidden layers, with 100 hidden neurons and a 'tanh' activation function. In both studies, the maximum iteration, i.e., 'max_itr', is set to 300, and the 'learning rate' is kept adaptive.
3. **Support Vector Machine (SVM):** The penalty term 'C' is used to avoid misclassification in the SVM. In the first part, we have set the penalty term 'C' to 180 and the kernel as 'linear' to classify data points. In the second part, we have set penalty term 'C' to 80 and the kernel as the Radial basis function, i.e., 'rbf'.
4. **Random Forest Ensemble (RFE):** It creates several decision trees, each constructed using a distinct bootstrap sample of the training dataset. The number of trees 'n' is a crucial hyperparameter for the random forest. Overfitting occurs when there are more trees. In the first part, we have set the number of trees 'n' to 10. While in the second part, we have set 'n' to 100.

6. Results and Discussion

In this section, we have discussed implementation and results obtained from our system. Separately, we presented automated grading of PowerPoint presentations and recognized the speaker's confidence from its audio. The final score of presentation quality is obtained by combining the output of the first and second parts of our system.

6.1. Evaluation of the PowerPoint presentations

Table 1, presented in section 4, shows the list of features used in our experimentation. Program is implemented in python to extract features from the PowerPoint presentations. Less important features are identified using the tree-based ExtraTree classifier approach. A randomly selected subset of the features creates multiple correlated decision trees, which helps find correlated features.

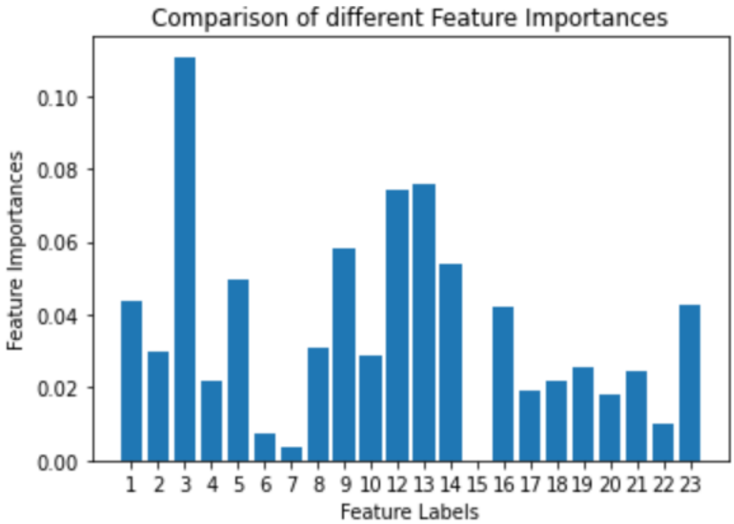

Figure 5: Features and their importance in the grading of the PowerPoint presentations

Feature labels in Figure 5 represent the feature number. A chart, slide number, and date are essential features because the average number of students have used these features. Features like placeholder, slide title, hyperlink, and font are least important because most students use or do not use these features in their PowerPoint presentations. These features are not helpful in our dataset, hence removed. Linear Discriminant Analysis is used to reduce the dimensionality of the features. Machine learning models implemented using SK-Learn library. These models are tested by feeding unseen PowerPoint presentations to them. The output of the machine learning model is categorized into 'Excellent', 'Very Good', 'Good', and 'Fair' classes.

Figure 6 depicts that the PowerPoint dataset is not 100% balanced. Class 'Good' has sufficient samples, while class 'Excellent' has a smaller number of samples. It indicates that fewer students have designed 'Excellent' PowerPoint presentations. Hence, our work will help students grade their PowerPoint presentations automatically and improve until they are satisfied with their grades.

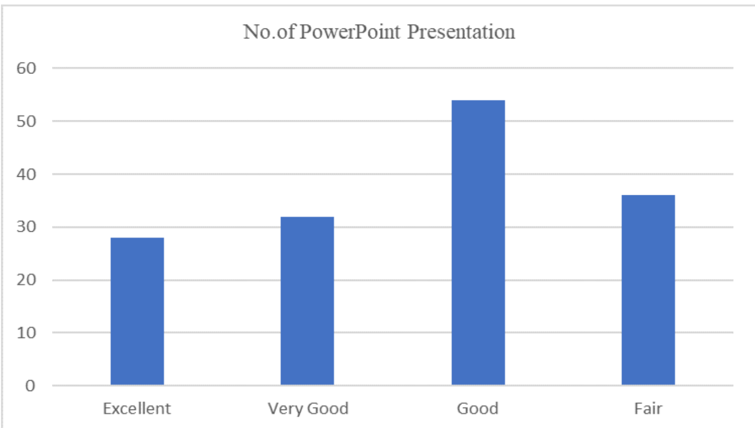

Figure 6: Dataset of the PowerPoint Presentations

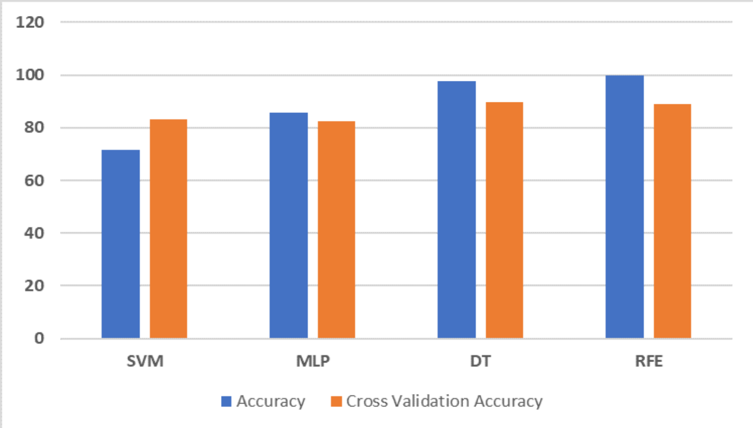

Figure 7: Classifier’s accuracy for grading of the PowerPoint presentations

Cross-Validation is a standard method for determining actual performance. In our experimentation, we have applied 6-fold cross-validation. Random Forest Ensemble gives 100% accuracy and 91.60% cross-validation accuracy in classifying the grade of PowerPoint presentations. The Decision Tree gives 97.61% accuracy and 89.74% cross-validation accuracy. MLP classifier has better accuracy compared to SVM. However, cross-validation accuracy is more in SVM than MLP. Figure 7 shows the accuracy of all the classifiers used in our experimentation.

Machine Learning Models	Evaluation Metrics	Output class of PowerPoint presentation			
		Excellent	Very Good	Good	Fair
Decision Tree	Precision	0.91	1.00	1.00	1.00
	Recall	1.00	0.88	1.00	1.00
	F1-score	0.95	0.93	1.00	1.00
Multilayer Perceptron	Precision	1.00	0.80	0.75	1.00
	Recall	1.00	1.00	0.86	0.67
	F1-score	1.00	0.89	0.80	0.80
Support Vector Machine	Precision	0.43	0.50	0.94	0.67
	Recall	0.75	0.43	0.73	0.89
	F1-score	0.55	0.46	0.82	0.76
Random Forest Ensemble	Precision	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	1.00	1.00
	F1-score	1.00	1.00	1.00	1.00

Table 2: Classifier’s Precision, Recall, and F1-score for grading of the PowerPoint presentations

Table 2 shows that the Random Forest Ensemble gives 1.00 Precision, Recall, and F1-score for all the classes. Results reveal that our system using a Random Forest Ensemble classifier gives correct results.

6.2. Evaluation of the confidence

We have extracted MFCC, Mel Spectrogram Frequency, and Chroma features using the Librosa python library. Linear Discriminant Analysis is used to reduce the dimensionality of the features. SK-learn python library is used to build machine learning models. Figure 8 shows the accuracy of all the classifiers used in our experimentations to recognize the speaker as 'Confident' or 'Non-Confident'. MLP has better accuracy compared to other classifiers. MLP gives 88.02% accuracy. RFE gives 84.90% accuracy. DT and SVM give 80.73% and 83.85% accuracy, respectively.

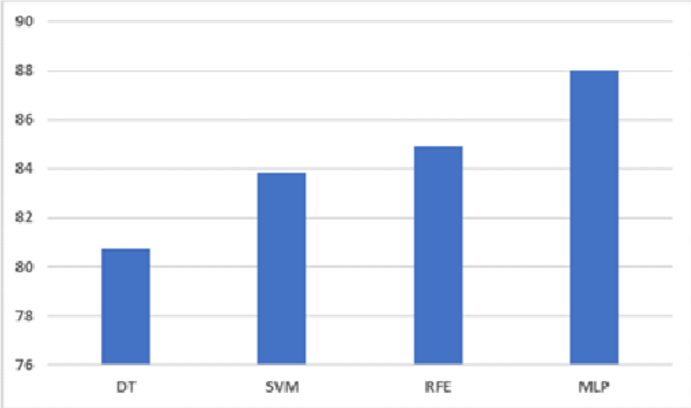

Figure 8: Classifier's accuracy for predicting the confidence of the students

Table 3 shows Precision, Recall, and F1-score for all the ML models used in our experimentation to classify students' audio into 'Confident' and 'Non-Confident' classes. Precision, Recall, and F1-score of MLP is more compared to other classifiers. MLP has performed better as compared to other classifiers. Results reveal that our system using MLP gives correct results.

Machine Learning Models	Evaluation Metrics	Output class of audio files	
		Confident	Non-Confident
Decision Tree	Precision	0.74	0.84
	Recall	0.83	0.76
	F1-score	0.78	0.80
Multilayer Perceptron	Precision	0.89	0.87
	Recall	0.84	0.91
	F1-score	0.87	0.89
Support Vector Machine	Precision	0.84	0.84
	Recall	0.81	0.87
	F1-score	0.82	0.85
Random Forest Ensemble	Precision	0.79	0.85
	Recall	0.83	0.82
	F1-score	0.81	0.83
Naïve Bayes	Precision	0.82	0.64
	Recall	0.38	0.93
	F1-score	0.52	0.76

Table 3: Classifier's Precision, Recall, and F1-score for predicting the confidence of the students

6.3. Evaluation of the Presentation skills

We have calculated the final score for the student's presentation skills by combining the score of PowerPoint presentations and the student's confidence level and compared the similarity between the final score given by our system and human evaluators. We have considered 20 sample sizes of labeled data, i.e., never exposed to our system. PowerPoint presentation and recorded audio of each student feed as input to our system to get a grade. Table 4 shows rubrics to assign a score to PowerPoint presentations and students' confidence depending on the grade.

PowerPoint presentation Grade				
Grade	Excellent	Very Good	Good	Fair
Marks	5	4	3	2
Confidence level of the speaker				
Grade	Confident		Non-Confident	
Marks	5		2	

Table 4: Rubrics to assign score depending on the grade

10 marks are allotted for the presentation skills. Our concern is only about grading the presentation skills. Out of 10 marks of presentation skills, 5 are for the PowerPoint presentation, and 5 for the confidence. Table 5 displays the score automatically calculated by our system and the score given by human raters.

Sr. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
AS	10	9	5	5	9	10	5	9	10	10	9	5	10	8	8	5	9	10	9	10
HRS	9	8	6	6	8	9	5	8	10	10	7	4	9	9	9	7	8	9	9	10

Table 5: Automated score (AS) versus Human rated score (HRS)

We have compared the final score allotted by human raters and our system using the Quadratic Weighted Kappa (QWK) evaluation metric. QWK value 1 indicates 100% similarity, 0 indicates no similarity, and 0.5 indicates a moderate similarity between the marks allotted by the human raters and our system. In our work, we have achieved a QWK value of 0.82, which is more than moderate similarity. It indicates that our system performs well in the automated grading of presentation skills.

7. Conclusions and Future Work

Our work adds to the existing knowledge area of automated grading. We have implemented a computer-based automated technique to assess students' presentation skills through PowerPoint presentations and audio. The teacher only needs to focus on the content or topic expertise of the students. Our system is also helpful to students, as it encourages the iterative refinement of the students' work before the actual presentation.

We have identified a set of valuable features from the PowerPoint presentations and audio by using Python−pptx and Librosa library, respectively. We have created a modest data set to allow the development of machine learning techniques. Unnecessary features are removed using the tree-based feature selection method. We have also employed the linear discriminant analysis technique to reduce dimensionality. Various prototype classifiers are used to demonstrate a data-driven method to assess presentation skills. The best classifier to grade PowerPoint slides is Random Forest Ensemble which gives 100% accuracy. The best classifier to grade confidence level is Multilayer Perceptron which gives 88.02 % accuracy.

We can extend our work of grading PowerPoint presentations by incorporating the technical contents of the presentations in the future. Seminar and Project reports play a crucial role in the termwork assessment, as they carry good credits. We can even further incorporate automatic grading of the report.

References

Agung Putri Ratna, A., D. Lalita Luhurkinanti, I. Ibrahim, D. Husna, and P. Dewi Purnamasari. 2018. "Automatic essay grading system for Japanese language examination using winnowing algorithm". In *Proceedings - 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life, iSemantic 2018*, 565-69. IEEE. https://doi.org/10.1109/ISEMANTIC.2018.8549789.

Ajitiono, T., and Y. Widyani. 2016. "Indonesian essay grading module using Natural Language Processing". In *Proceedings of 2016 International Conference on Data and Software Engineering, ICoDSE 2016*, 1-5. IEEE. https://doi.org/10.1109/ICODSE.2016.7936117.

Al-Jouie, M. F., and A. M. Azmi. 2017. "Automated evaluation of school children essays in Arabic". *Procedia Computer Science* 117: 19-22. https://doi.org/10.1016/j.procs.2017.10.089.

Azmi, A. M., M. F. Al-Jouie, and M. Hussain. 2019. "AAEE – Automated evaluation of students' essays in Arabic language". *Information Processing and Management* 56, no. 5: 1736-52. https://doi.org/10.1016/j.ipm.2019.05.008.

Bashir, A. M., A. Hassan, B. Rosman, D. Duma, and M. Ahmed. 2018. "Implementation of a neural natural language understanding component for Arabic dialogue systems". *Procedia Computer Science* 142: 222-29. https://doi.org/10.1016/j.procs.2018.10.479.

Borade, J. G., A. W. Kiwelekar, and L. D. Netak. 2022. "Feature extraction for automatic grading of students' presentations". In *ICT Systems and Sustainability*, 293-301. Lecture Notes in Networks and Systems, Vol. 321. https://doi.org/10.1007/978-981-16-5987-4_30.

Borade, J. G., and L. D. Netak. 2021. "Automated grading of essays: A Review". In *Intelligent Human Computer Interaction. IHCI 2020*, 238-49. Lecture Notes in Computer Science, Vol. 12615. Springer International Publishing. https://doi.org/10.1007/978-3-030-68449-5_25.

Dar, S. A., and Z. Khaki. 2013. "Emotion recognition based on audio speech". *IOSR Journal of Computer Engineering (IOSR-JCE)* 11, no. 6: 46-50.

Dat, P. T., and L. The Anh. 2019. "Application of convolutional neural network for gender and age group recognition from speech". In *Proceedings - 2019 6th NAFOSTED Conference on Information and Computer Science, NICS 2019*, 489-93. IEEE. https://doi.org/10.1109/NICS48868.2019.9023870.

Deshmukh, G., A. Gaonkar, G. Golwalkar, and S. Kulkarni. 2019. "Speech based emotion recognition using machine learning". In *Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019*, 812-17. IEEE. https://doi.org/10.1109/ICCMC.2019.8819858.

Fazal, A., F. K. Hussain, and T. S. Dillon. 2013. "An innovative approach for automatically grading spelling in essays using rubric-based scoring". *Journal of Computer and System Sciences* 79: 1040-56. https://doi.org/10.1016/j.jcss.2013.01.021.

George, N., P. J. Sijimol, and S. M. Varghese. 2019. "Grading descriptive answer scripts using deep learning". *International Journal of Innovative Technology and Exploring Engineering* 8, no. 5: 991-96.

Harb, H., and L. Chen. 2003. "Gender identification using a general audio classifier". In *Proceedings - IEEE International Conference on Multimedia and Expo*, II733-II36. IEEE. https://doi.org/10.1109/ICME.2003.1221721.

Janda, H. K., A. Pawar, S. Du, and V. Mago. 2019. "Syntactic, semantic and sentiment analysis: The joint effect on automated essay evaluation". *IEEE Access* 7: 108486-503. https://doi.org/10.1109/ACCESS.2019.2933354.

Kattel, M., A. Nepal, A. K. Shah, and D. Shrestha. 2019. "Chroma feature extraction". In *Conference: Chroma Feature Extraction using Fourier Transform*.

Kuchebo, A. V., V. V. Bazanov, I. Kondratev, and A. M. Kataeva. 2021. "Convolution neural network efficiency research in gender and age classification from speech". In *Proceedings of the 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2021*, 2145-49. IEEE. https://doi.org/10.1109/ElConRus51938.2021.9396365.

Ladde, P. P., and V. S. Deshmukh. 2015. "Use of multiple classifier system for gender driven speech emotion recognition". In *Proceedings - 2015 International Conference on Computational Intelligence and Communication Networks, CICN 2015*, 713-17. IEEE. https://doi.org/10.1109/CICN.2015.145.

Leng, Y., L. Yu, and J. Xiong. 2019. "DeepReviewer: Collaborative grammar and innovation neural network for automatic paper review". In *ICMI 2019 - Proceedings of the 2019 International Conference on Multimodal Interaction*, 395-403. https://doi.org/10.1145/3340555.3353766.

Parihar, S., R. Das, Z. Dadachanji, A. Karkare, P. K. Singh, and A. Bhattacharya. 2017. "Automatic grading and feedback using program repair for introductory programming courses". In *ITiCSE '17: Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*, 92-97. https://doi.org/10.1145/3059009.3059026.

Peng, X., D. Ke, Z. Chen, and B. Xu. 2010. "Automated chinese essay scoring using vector space models". In *Proceedings of IUCS 2010 - 2010 4th International Universal Communication Symposium*, 149-53. IEEE. https://doi.org/10.1109/IUCS.2010.5666229.

Surya, K., E. Gayakwad, and M. K. Nallakaruppan. 2019. "Deep learning for short answer scoring". *International Journal of Recent Technology and Engineering* 7, no. 6: 1712-15.

Voigtlaender, P., P. Doetsch, and H. Ney. 2016. "Handwriting recognition with large multidimensional long short-term memory recurrent neural networks". In *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR*, 228-33. IEEE. https://doi.org/10.1109/ICFHR.2016.0052.

Walia, T. S., G. S. Josan, and A. Singh. 2019. "An efficient automated answer scoring system for Punjabi language". *Egyptian Informatics Journal* 20, no. 2: 89-96. https://doi.org/10.1016/j.eij.2018.11.001.

Yucesoy, E., and V. V. Nabiyev. 2013. "Gender identification of a speaker using MFCC and GMM". In *ELECO 2013 - 8th International Conference on Electrical and Electronics Engineering*, 626-29. IEEE. https://doi.org/10.1109/eleco.2013.6713922.

Zeng, Y. M., Z. Y. Wu, T. Falk, and W. Y. Chan. 2006. "Robust GMM based gender classification using pitch and RASTA-PLP parameters of speech". In *Proceedings of the 2006 International Conference on Machine Learning and Cybernetics*, 3376-79. IEEE. https://doi.org/10.1109/ICMLC.2006.258497.