IntelliView: An AI-Mock Interview Platform

Swati Uparkar[1], Saurabh Hundare[2], Varun Gazala[3] , Sarvesh Chaudhari[4] , Ankush Jain[5]

[1]*Artificial Intelligence and Data Science Department*

[2]*Artificial Intelligence and Data Science Department*

[3]*Artificial Intelligence and Data Science Department*

[4]*Artificial Intelligence and Data Science Department*

[5]*Artificial Intelligence and Data Science Department*

---***---

Abstract - IntelliView is a groundbreaking platform that aids novice job seekers by harnessing the power of advanced AI and NLP technologies. It combines HTML, CSS, JavaScript, and the Deep Face method to offer a comprehensive real-time interview practice experience. Users can engage in text and video-based interviews and receive detailed feedback on their verbal and non-verbal communication skills. Additionally, the platform provides an audio analysis module that utilizes CNN-LSTM to evaluate emotional states during interviews, offering valuable insights for improvement. Users can also take advantage of IntelliView's resume builder to create personalized, professional resumes that cater to specific industry standards. The platform's seamless integration of AI and NLP technologies helps guide first-time job seekers in refining their communication techniques and understanding their emotional responses. By offering these transformative tools, IntelliView empowers users to approach the competitive job market with confidence and competence.

Key Words: Interview Assessment, DeepFace, NLP, Text-based analysis, Video-based analysis, Resume Builder, Audio-based analysis

1.INTRODUCTION

The combination of artificial intelligence (AI) and natural language processing (NLP) has produced major advances in interview evaluation and job application preparation in the fast-paced world of contemporary employment. This study examines the IntelliView project, a creative endeavor that integrates cutting-edge technology to assist job seekers seeking entry-level positions. There are four key components in the IntelliView platform that cover different facets of the job search process. Through the use of real-time interview scenarios, the text-based analysis module assesses users' responses based on predetermined standards. This gives valuable comments to improve overall performance in the interview as well as a realistic interview practice setting.

An important change is the addition of the video-based analysis module, which goes beyond typical examinations. This module uses the Deep Face technique to do sentiment and emotion analysis in real time, giving users important insights into the dynamics of their nonverbal communication during interviews. Candidates can refine their emotional intelligence with the use of such feedback, which enables them to modify their communication tactics accordingly.

The third IntelliView module is then a powerful resume creator that is carefully crafted using HTML, CSS, and JavaScript. This feature gives job seekers a complete toolkit for creating custom resumes by providing a variety of templates that can be adjusted to meet specific demands. When taken as a whole, these modules represent a paradigm change in the field of job preparation by balancing advanced technology with subtle human contact.

Finally, the fourth module of IntelliView functions as an audio-based analysis tool that employs CNN-LSTM to evaluate users' voices and emotional tones during interviews. This feedback helps users refine their speaking style, respond to questions with confidence, and maintain a balanced tone throughout conversations.

2. LITERATURE SURVEY

Here is a survey of pertinent literature techniques. It outlines the several methods that were employed. The brief information about the referred research papers is explained in this section.

In Paper[1], the authors investigate the feasibility of using chest X-rays as a screening tool for COVID-19 in regions facing testing kit shortages. By applying deep convolutional neural networks, the study achieves a high classification accuracy of 90.64% and an F1-Score of 89.8%, suggesting the potential for X-ray-based diagnosis as an alternative approach in resource-constrained settings.

In Paper[2], it explains how to utilize a software program that uses the Haar-Cascade Algorithm with a pre-trained model called DeepFace to identify various human emotions.

In Paper[3], the research's study objective is to gauge how semantically equivalent multi word sentences are for the guidelines and procedures found in railway safety documentation.

In Paper[4], the objective of this paper was to create an interview simulation using Deep learning and speech-to- text systems.

In Paper[5], it delves into the analysis of emotion detection and places a focus on blink count as well.

In Paper[6], it primarily centers on the utilization of NLP techniques, specifically NLTK and Ngrams, for assessing text similarity. It serves as a guide for conducting text similarity analysis on provided input.

In Paper[7], it introduces a real-time facial emotion detection system using OpenCV, Deep-Face, and TensorFlow. It aids narcotics officers in suspect identification and assists robots in recognizing emotions like happiness, nervousness, and neutrality.

In Paper[8], the paper presents a novel Face- Based Video Retrieval (FBVR) pipeline designed for unconstrained television- like videos. It introduces a new dataset for evaluation and achieves high retrieval accuracy (97.25% mean average precision) while maintaining real-time processing speed.

In Paper[9], the suggested algorithm in this study automatically evaluates and forecasts an interviewee's nonverbal cues and offers pertinent comments.

In Paper[10], the mock-interview platform that is suggested in this study assesses candidates' traits of personality and interview performance in addition to analyzing the textual, audio, and visual aspects of an interview.

In Paper [11], the research provides insights into the effectiveness of combining 3D CNN and LSTM networks with a self-attention mechanism for audio analysis, specifically in recognizing human emotions. Utilizing spectrograms as input and applying a relation-aware approach, the model demonstrates improved accuracy in speech emotion recognition. The approach offers enhanced feature extraction and sequence-to-sequence parallelization, leading to better performance in the task.

In Paper [12], the objective of this paper is to perform a comparative analysis of different datasets in the context of audio analysis using the CNN-LSTM method. This approach seeks to evaluate how the model performs across diverse data samples, allowing for insights into its generalizability and effectiveness in different audio recognition scenarios.

3. PROPOSED WORK

This research paper explores a comprehensive approach that combines emotion detection, text analysis, and resume building to leverage textual data effectively. Text Analysis methods, including Natural Language Processing (NLP) and Machine Learning Algorithms, are used to analyze and compare similarities between the actual and expected solutions. Furthermore, the paper discusses the application of these techniques in resume building where HTML, CSS and JavaScript is used.

Video Analysis

For video analysis facial recognition is a really an important step, for which we have used Haar Cascade face detection. Haar cascade face detection is a method used for detecting faces in images or video. It is based on the Haar wavelet technique and is an effective and computationally efficient way to perform face detection. The Haar Cascade face detection is widely used due to its simplicity and efficiency.

DeepFace

Architecture: Max-pooling, fully linked, and convolutional layers are present in various layers of DeepFace. It utilizes a 3D face model to map faces into a 3D space, allowing for pose-invariant face recognition.

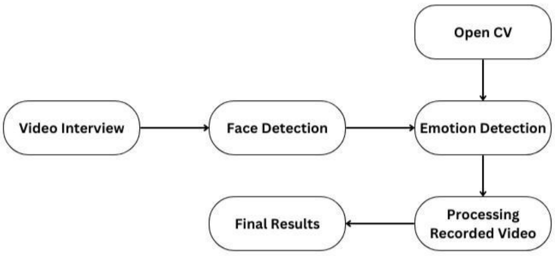
Face Alignment: DeepFace aligns faces to a canonical pose, reducing variations caused by pose and improving recognition accuracy.

Metric Learning: It uses a metric learning approach to learn a similarity function that compares faces in a continuous representation space.

CNN Overview

Basic Architecture: Convolutional, pooling, and fully linked layers are the components of CNNs. While pooling layers decrease the spatial dimensions of the features, convolutional layers extract features from input images.

Training: Backpropagation and gradient descent are commonly used in CNN training to minimize a loss function, including mean squared error or cross-entropy loss.



Video Interview Flow

B. Text Analysis

For the project, the system takes input from the users and then processes it further to generate outcomes and similarities.

Text analysis encompasses a wide range of techniques aimed at understanding, interpreting, and extracting meaningful information from text. At its core, text analysis involves processing and analyzing textual data to uncover patterns, trends, and insights that can inform decision-making, drive innovation, and enhance understanding.

Here's a simple breakdown of how it works:

Preprocessing: Before calculating similarity, the input sentences are preprocessed. This involves:

Tokenization: Breaking down each sentence into individual words or tokens.

Lowercasing: To maintain uniformity, all tokens will be converted to lowercase.

Removing stopwords: Words that are too common to be valuable for similarity comparison, like 'and', 'the', etc., are removed.

Stemming: word reduction to its root (e.g., "walking" becomes "walk").

TF-IDF Vectorization: Once the preprocessing is done, the preprocessed sentences are transformed into numerical vectors using the TF-IDF vectorizer. The TF-IDF measures a word's significance within a sentence in relation to a corpus, or group of sentences.

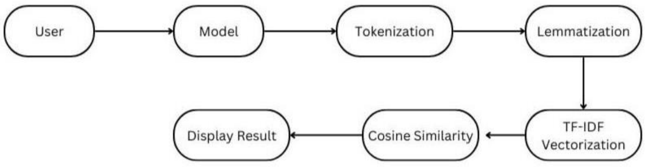
Cosine Similarity Calculation: The cosine similarity between the two texts is derived post vectorization. No matter how big or small the angle between two vectors is, cosine similarity calculates the cosine of that angle to show how similar they are. Higher cosine similarity indicates higher similarity between the sentences.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2}\sqrt{\sum\limits_{i=1}^{n} B_i^2}},$$

Formula for Cosine Similarity

Return: The similarity score, which goes from 0 to 1—0 denoting no resemblance and 1 denoting sentence similarity—is returned by the function.

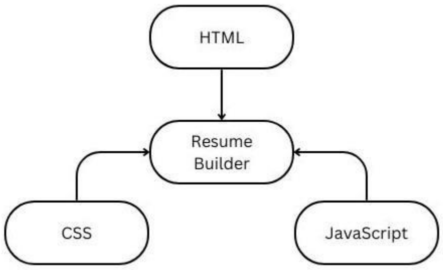
So, in simple terms, this function takes two sentences, converts them into numerical representations using TF-IDF, then calculates the cosine similarity between these representations to measure how similar the sentences are.



Text Analysis Flow

C. Resume Builder

The digital age has transformed the job application process, with online resumes becoming a standard requirement. To meet this demand, individuals can now create dynamic and visually appealing resumes using web technologies such as HTML, CSS, and JavaScript. This research paper presents a comprehensive guide to building an interactive resume builder application using these technologies. The paper covers key aspects of resume building, including form design, data validation, dynamic preview, and PDF generation.This research paper provides a practical guide for building an interactive resume builder application using HTML, CSS, and JavaScript.



Resume Builder Flow

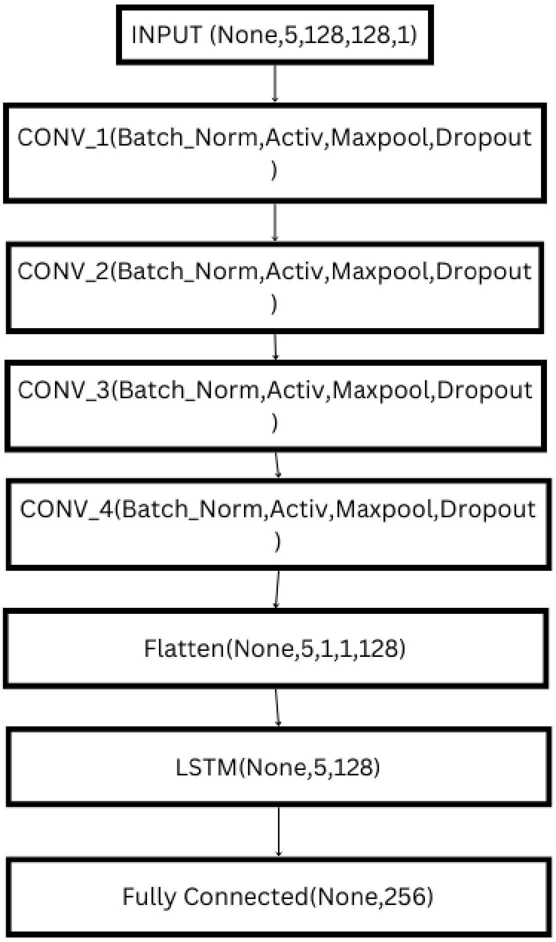
4.Audio Analysis

For audio analysis, we have created a Time Distributed CNN-LSTM model using the RAVDESS dataset for training.

A. CNN-LSTM model

To recognize sequential patterns in audio signals, the Time distributed convolutional neural network combines an LSTM-based recurrent neural network with hierarchical CNNs.

It uses a rolling window approach and operates directly on log-mel spectrograms. It processes each segment using a convolutional neural network that has four Local Feature Learning Blocks (LFLBs). In order to capture long-term dependencies, the output is then input into a recurrent neural network that has two LSTM cells. This results in a fully connected layer with softmax activation for emotion prediction.


INPUT (None,5,128,128,1)
CONV_1(Batch_Norm,Activ,Maxpool,Dropout)
CONV_2(Batch_Norm,Activ,Maxpool,Dropout)
CONV_3(Batch_Norm,Activ,Maxpool,Dropout)
CONV_4(Batch_Norm,Activ,Maxpool,Dropout)
Flatten(None,5,1,1,128)
LSTM(None,5,128)
Fully Connected(None,256)

Audio Analysis Neural Network

4. METHODOLOGY

To recognize sequential patterns in audio signals, the Time To effectively leverage textual data in employment preparation, the proposed work takes a multifaceted approach integrating resume building, text analysis, and emotion detection. In order to put this all-encompassing strategy into practice, a methodical methodology is developed that consists of discrete steps that are specific to each project component.

First, the methodology starts with gathering and preparing textual data from multiple sources, such as external datasets and user inputs. The textual data is preprocessed using natural language processing (NLP) techniques, which include tokenization, lowercasing, stopword removal, and stemming. By ensuring the textual data is consistent and tidy, this pretreatment phase makes reliable analysis and comparison possible.

The methodology goes into the construction of text analysis and emotion detection modules after data initial treatment. In order to identify emotions in video analysis, faces in video streams are recognized using the Haar cascade face detection technique. Next, facial expressions and feelings are analyzed using DeepFace architecture, which aligns faces to a canonical pose and uses metric learning to compare faces in a continuous representation space. Text analysis methods are used in parallel to examine and contrast user-provided textual inputs. Preprocessed sentences are vectorized using the TF-IDF method, and the similarity between sentences is then calculated using the cosine similarity approach.

In parallel, the methodology proceeds with the development of the resume builder module, leveraging web technologies such as HTML, CSS, and JavaScript. The resume builder application is designed to offer users a dynamic and visually appealing platform for crafting personalized resumes. Key aspects of resume building, including form design, data validation, dynamic preview, and PDF generation, are meticulously implemented to enhance user experience and functionality.

The methodology also includes the audio emotion detection module which detects the emotion of the input audio file. It is created by the training using the audio data and the process of data augmentation to enhance the robustness of the module. After the CNN-LSTM model is trained which is the Time Distributed CNN model that divides the audio files in spectrograms and identifies the most dominating emotion using LSTM for long term dependencies and fully connected layer and softmax activation function.

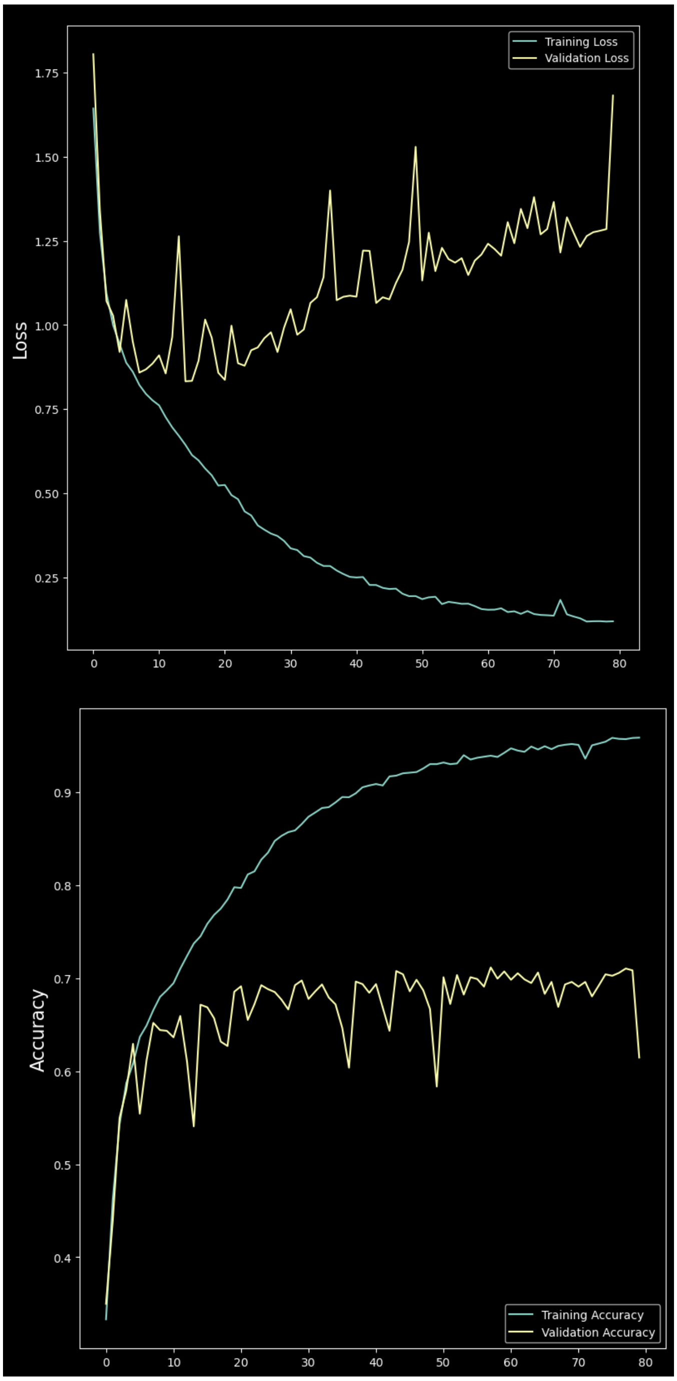
Iterative testing and validation are carried out during the implementation phase to guarantee the reliability and efficiency of every module. In order to determine areas that require improvement and refinement, user feedback and performance metrics are gathered and examined. The process is based on constant iteration and optimization, guaranteeing that IntelliView, the finished product, is an excellent example of technology used wisely for job preparation.

5. PERFORMANCE EVALUATION

We compared the performance of two models in the first stage of the video interview project: the pre-trained DeepFace model and our personally designed CNN model. These tests were carried out in the Jupyter Notebook environment.

The Metrics used for the proprietary trained CNN model are loss:- ‘Categorical CrossEntropy’, Learning Rate:- 0.0005, Epochs:- 80.

- Custom Trained Model (CNN):
  Loss:- 0.1197
  Accuracy:- 0.94
  Validation_Accuracy:- 0.62
  Validation_Loss:- 1.6816



As of in this graph we can see the accuracy vs the loss graph which concludes upto the assumption that model has slightly overfitted due to large number of epochs.

We tried to reduce the number of epochs but could not match the accuracy of the Deepface model as its accuracy was 97% and as a result used the Deepface model as our final model for the Video Based Interview process.

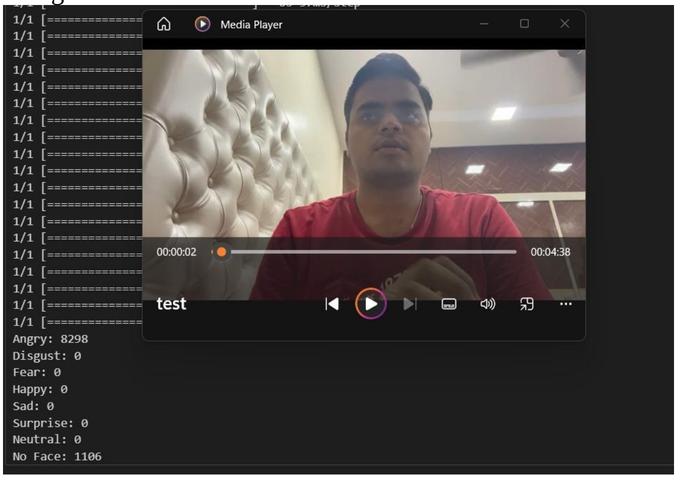
The testing of the both models are depicted in the further section of the performance evaluation.

Testing of Both the models are done using the same video recording of a candidate answering to the questions that are presented on the screen.

1. Result of the Custom Trained CNN model:

In this the model is fed with a Video of the candidate giving answers to particular questions and as we can see that the model divides each frame of the video in order to capture the Emotion from each frame for the final result.

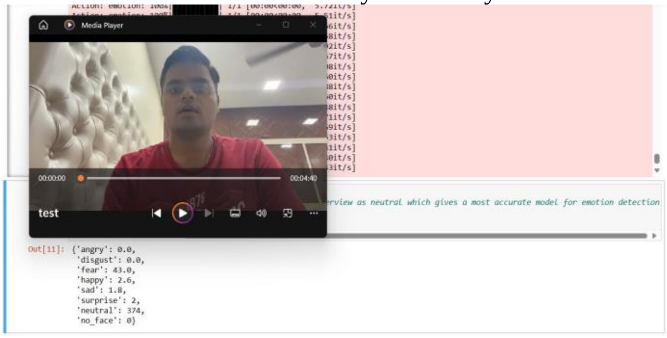
As we can see that model is not displaying satisfactory results though it has been overfitted.



```
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
1/1 [=============
Angry: 8298
Disgust: 0
Fear: 0
Happy: 0
Sad: 0
Surprise: 0
Neutral: 0
No Face: 1106
```

2. Result of the DeepFace model:

On the other hand DeepFace model also operates on the same principle of dividing each frame of the video and processing it and the results are also satisfactory with the system needs.



```
erview as neutral which gives a most accurate model for emotion detection

Out[11]: {'angry': 0.0,
          'disgust': 0.0,
          'fear': 43.0,
          'happy': 2.6,
          'sad': 1.8,
          'surprise': 2,
          'neutral': 374,
          'no_face': 0}
```

With an incredible precision score of 97.2%, the Deepface Model is the obvious choice for our application.

In summary, the key performance evaluation parameters for our model are model selection (Deepface Model), an accuracy rate of 97.2%, and our application preference for the Deepface Model due to its higher accuracy and graphical representation capabilities. This choice ensures that our application will provide users with the most accurate and visually appealing results.

The Second part of the project focuses on the Text Based Interview model that depicts the testing of the candidate based on the textual Question-Answering process.

For this we have tried two models that are the pre-trained model of spacy 'en_core_web_sm' and a custom nlp model that particularly uses cosine similarity.

Pre-trained en_core_web_sm model:

In this we just imported the model using the spacy library. The model is a starter kit for the text processing although giving a satisfactory accuracy for the given task.

Custom NLP model:

Particularly in this we have manually performed the pipeline of natural language processing which includes tokenization followed by stemming which is employed by PorterStemmer.

For calculating the similarity of the user text and the actual answer the similarity method used are Jaccard Similarity and Cosine Similarity.

Jaccard Similarity:

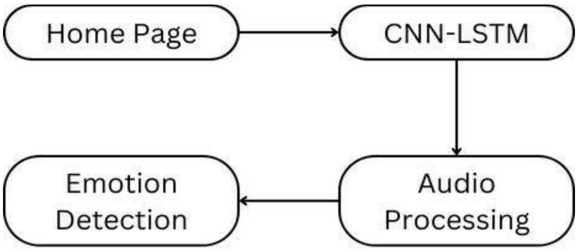
While employing this method gave promising results but lacked in giving more accurate results.

Cosine Similarity:

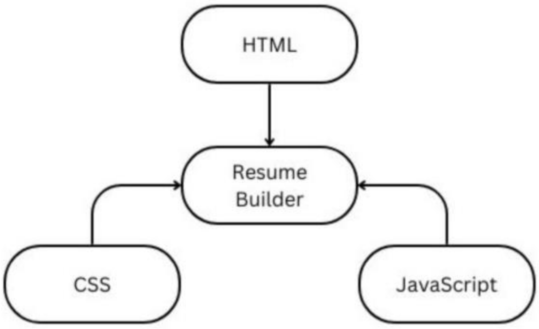
This method has given us the highest accuracy due to its ability to compare word vectors or embeddings, which represent the multi-dimensional meanings of words.
The key benefit of cosine similarity is that it works well for text document comparisons and can handle big documents.
As a result, the Custom NLP model which uses cosine similarity has been employed as it gives good accuracy for the text similarity task.

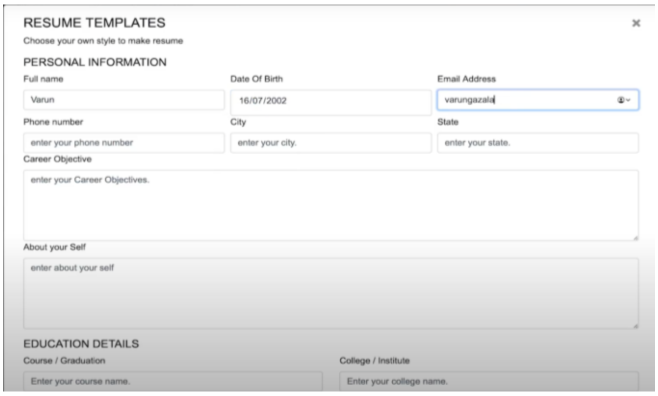
The third part of the project focuses on the Audio Based Assessment model that depicts the testing of the candidate based on the textual Question-Answering process.

For this we have used the CNN-LSTM model.



The fourth part of the project focuses on the Resume Builder model that uses the frontend technologies to build a professional resume. You have to just enter the details asked and then automatically a resume is generated which is static and downloadable in pdf format.





RESUME TEMPLATES
Choose your own style to make resume

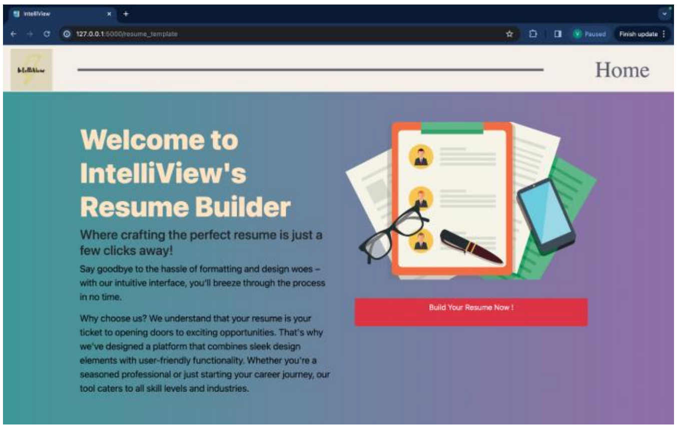
PERSONAL INFORMATION

Full name: Varun
Date Of Birth: 16/07/2002
Email Address: varungazala
Phone number: enter your phone number
City: enter your city.
State: enter your state.
Career Objective: enter your Career Objectives.
About your Self: enter about your self

EDUCATION DETAILS

Course / Graduation: Enter your course name.
College / Institute: Enter your college name.

This is the input page where you have to enter the details.



Welcome to IntelliView's Resume Builder

Where crafting the perfect resume is just a few clicks away!

Say goodbye to the hassle of formatting and design woes – with our intuitive interface, you'll breeze through the process in no time.

Why choose us? We understand that your resume is your ticket to opening doors to exciting opportunities. That's why we've designed a platform that combines sleek design elements with user-friendly functionality. Whether you're a seasoned professional or just starting your career journey, our tool caters to all skill levels and industries.

This is the homepage where you can read about our resume builder model.

6. CONCLUSION

To sum up, the IntelliView project is revolutionizing the process of preparing job applications by utilizing cutting edge technologies to enable job seekers at the entry level. Through a comprehensive framework integrating AI, NLP, and the innovative Deep Face method, IntelliView redefines interview analysis with real-time assessments, constructive feedback, and dynamic insights into non-verbal communication. The CNN-LSTM is used to effectively improve your tone and confidence and provides a way to speak properly at the time of a real interview. The multifaceted modules collectively signify a paradigm shift, providing a nuanced approach to the competitive job market. As technology and human interaction converge, IntelliView equips individuals with essential competencies, offering a glimpse into the future of job application preparation. This research underscores the project's technological underpinnings, methodological rigor, and its broader implications for reshaping the landscape of employment preparation.

7. FUTURE SCOPE

The future scope of the IntelliView project is promising, poised to witness continued evolution and expansion in response to the dynamic landscape of job application preparation. Potential avenues for enhancement include the incorporation of additional AI algorithms to diversify the range of interview scenarios and augment the accuracy of feedback mechanisms.

Further exploration into personalized learning algorithms could tailor the platform to individual user needs, ensuring a more adaptive and user-centric experience. Additionally, ongoing developments in natural language understanding and emotion recognition technologies could refine the project's capabilities in assessing both verbal and non-verbal communication. Collaborations with industry stakeholders and academia may facilitate the integration of real-world job market insights, fostering a more comprehensive and relevant tool for job seekers. The continual integration of emerging technologies and iterative refinement will be pivotal in sustaining the IntelliView project's efficacy and relevance in the ever-evolving landscape of employment preparation.

REFERENCES

1. Jadhav, Aaditya & Ghodake, Rushikesh & Muralidharan, Karthik & Varma, G & Jagan, Vijaya Bharathi. (2023). INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM) AI Based Multimodal Emotion and Behavior Analysis of Interviewee. 10.55041/IJSREM19049.

2. J. Kaur, J. Saxena, J. Shah, Fahad and S. P. Yadav, "Facial Emotion Recognition," 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2022, pp. 528-533, doi: 10.1109/CISES54857.2022.9844366.

3. A. W. Qurashi, V. Holmes and A. P. Johnson, "Document Processing: Methods for Semantic Text Similarity Analysis," 2020 International Conference on Innovations in Intelligent Systems and Applications (INISTA), Novi Sad, Serbia, 2020, pp. 1-6, doi: 10.1109/INISTA49547.2020.9194665.

4. Sahil Temgire , Akash Butte , Rohan Patil , Varun Nanekar, Shivganga Gavhane, 2021, Real Time Mock Interview using Deep Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 05 (May 2021), doi : 10.17577/IJERTV10IS050213

5. A. K. A, A. H, N. P. Nair, V. A and A. T, "Interview Performance Analysis using Emotion Detection," 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2022, pp. 1424-1427, doi: 10.1109/ICIRCA54612.2022.9985667.

6. S. K. Sinha, S. Yadav and B. Verma, "NLP-based Automatic Answer Evaluation," 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode,India, 2022, pp. 807-811, doi: 10.1109/ICCMC53470.2022.9754052.

7. N. C. Brintha, J. A. Narayana, G. L. V. S. Jaswanth, G. J. Chandrapal and D. Venkat, "Realtime Facial Emotion Detection Using Machine Learning," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2022, pp. 1-5, doi: 10.1109/ICSES55317.2022.9914318.

8. Ciaparrone, G., Chiariglione, L. & Tagliaferri, R. A comparison of deep learning models for end-to-end face-based video retrieval in unconstrained videos. Neural Comput & Applic 34,7489–7506 (2022) https://doi.org/10.1007/s00521-021-06875-x

9. M. S. P, D. Hepsi Priya, P. Malavika and L. A, "Automated Analysis and Behavioural Prediction of Interview Performance using Computer Vision," 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1-6, doi: 10.1109/INDICON56171.2022.10039785.

10. Y. -C. Chou, F. R. Wongso, C. -Y. Chao and H. -Y. Yu, "An AI Mock-interview Platform for Interview Performance Analysis," 2022 10th International Conference on Information and Education Technology (ICIET), Matsue, Japan, 2022, pp. 37-41, doi:10.1109/ICIET55102.2022.9778999.

11. Neha Prerna Tigga, & Shruti Garg. (2023). Speech Emotion Recognition for multiclass classification using Hybrid CNN-LSTM. International Journal of Microsystems and Iot, 1(1), 9–17. https://doi.org/10.5281/zenodo.8158288

12. Dangol, R., Alsadoon, A., Prasad, P.W.C. et al. Speech Emotion Recognition UsingConvolutional Neural Network and Long-Short TermMemory. Multimed Tools Appl 79, 32917–32934 (2020). https://doi.org/10.1007/s11042-020-09693-w

13. Swati Uparkar, Saurabh Hundare, Varun Gazala, Sarvesh Chaudhari, Ankush Jain. “IntelliView: An AI Based Mock Interview Platform”, 2024 International Journal of Scientific Research in Engineering and Management (IJSREM), https://www.doi.org/10.55041/IJSREM29201.