جامعة حلوان
HELWAN UNIVERSITY

Intelligent Video Interview

A graduation project dissertation by:

Alaa Mohamed Abdou
Hagar Mostafa
Mona Sayed Shazli
Hadeer Mohamed
Marina Nabil

Submitted in: 23/17/2021

Supervised By:

DR. Amr Al-Sayed

2020

Members’ contribution:

Student Name	Building code	Book
-Alaa mohamed -Hagar mostafa	• Emotion recognition • Detect filler words • Speech to text	**Chapter 2** Literature Review on Emotion Analysis **Chapter 6** Experiments, Results and Discussion **Parts from chapter 4** 4.1 Facial Expression Emotion Analysis
-Mona sayed -Hadeer mohamed	• Lie detection blinking pursedLipsDetector • Building report	**Chapter 3** Literature Review on Lie Detection **Chapter 7** Conclusion and Future work **Parts from chapter 4** 4.3 Anxiety Detection
-Marina Nabil	• Responsible for building the platform • Design all front pages that appear to interviewer and interviewee	**Chapter 1** Introduction **Chapter 5** Smart Interview Platform **Parts from chapter 4** 4.5 Platform Statistical Review Analysis

Abstract

The Internet is widely being used to make our lives easier and the world a smaller place, on this basis, online interviews are gaining more popularity. That requires an interviewer to decide whether or not an interviewee matches the criteria of a job description. Intelligent video interview (IVI) seeks to automate matching an interviewee by answering the interview's questions on a video, the feedback of our platform will give the information relevant to the interviewer to decide whether or not an interviewee is good.

Intelligent video interview (IVI) divided into emotion recognition analysis, lie detection powered by cognitive techniques and speech analysis to give relevant information to be analyzed by the interviewer.

Our proposed approach facilitates handling hundreds of interviewees by combining simple machine learning tasks together to form a relevant report for each interviewee, traditional analysis of hundreds of interviewees takes months where our platform (IVI) handles them within a day or two.

The amount of interviewees applying for a certain job in big companies are relatively big and such a platform will reduce time and costs for these companies.

Chapter 1

Introduction

1.1 Overview

The estimates for the 2019-2020 academic year indicate a total of 3,898,000 college graduates in the United States. This represents an increase of nearly six million students from the 2015-2016 academic year in the United States alone all of them need to find a job, and not just graduate apply for jobs, juniors, seniors, managers and all the hierarchy at one point will get interviewed to rank up in their own fields.

Companies are now being overwhelmed by the amount of resumes and emails that are sent to them on a daily basis, which puts a lot of effort on the hiring manager to check the email, categorize the emails, open them, check the cover letters and the CVs and shortlisting them based on the job and company requirements. Even if the hiring manager does an excellent job doing all the mentioned tasks. His decisions to choose a candidate for the interview will still be biased toward the first (or the last) few emails that have been sent. Which is unfair for the candidates who are obviously more fitting into the job and not the best for the companies that are trying to find the most fitting employees.

Since online interviews are gaining more popularity, the chances of all these people to apply for a job contain an online interview is huge, and to facilitate and ease the amount of work an interviewer has to do which include examining the interviewee and extracting whether or not he is qualified for the applying job on hundreds of interviewees, he can simply use our platform for filtering interviewees.

IVI filtering process is based on questions determined by the interviewer and the interviewee answers them in a qualified manner by recording himself through our website, as soon as he submit his recording, the video will be processed to extract the sequence of emotions, detecting whether or not the interviewee is anxious and extracting text from the interviewee's speech for the interviewee for skill extraction using his speech and his resume.

IVI can be a huge benefit in cost and time reduction for any major companies with hundreds of interviews done within its system.

IVI reduces the amount of work an interviewer has to do to find a good candidate for a job criteria while increasing the chances of filtering bad candidates in early stages.

1.2 Research Problem

While manually examining an interviewee's answer to an interview question is easy but time consuming, configuring our platform to extract useful information is hard to formulate and contains certain steps in a certain order, thus, the use of steps without physiological order may be false.

1.3 Research Motivation

There is a need for an interactive platform to automate filtering interviewee, however extracting candidate features from interviewees and effectively communicating this remains a challenge.

Whether it is to measure the effectiveness of the candidates, reduce the interview's costs, improve the interviewer performance, our platform while being a great help to major companies.

1.4 Research Objectives

One key area for companies and business intelligence is revenue prediction. One means of revenue prediction is utilizing all companies' functions to run on the highest performance and lowest cost. Currently, very few tools exist that effectively enable the utilization of automated interviews.

Organizations are faced with the need to hire extremely good interviewers in order to provide the company with the best candidate, but with hundreds of interviewees the chance of a slip is relatively high.

Our objective of this platform are to present new approaches for enhancing the interview's candidate selection and automate most of the process over our website for online interviews to achieve the following business benefits: automate most of the interview process, reducing the chance of an interviewer mistakes, reduce the work done by the interviewer which in return reduce the cost and utilize time more effectively.

And proposing an abstractive candidate selection summarizer that could visually represent a candidate is good or not, and showing the candidate skills.

1.5 System Requirements

Interviewer	• Sign up form for interviewer or company include:- 1. Company name 2. Email 3. Password 4. Number 5. Website • The interviewer can : Create the questions that the interviewer wants to put Determine the number of questions Determine the time for each question • The interviewer can : Make invitations for candidates Choose deadline for each invitation Put the name and email for each interviewee
candidate	The interviewee should: • Be direct to the page of interview after clicking on the invitation link • Assign his name and his email to Begin the interview • Submit All recorded questions

1.6 Stake Holders

- Companies that want to employ people (this company is represented by a member /hr./interviewer or owner

- Applicants (interviewee)applying for a job in a specific company are using this site for recruitment

1.7 Actors and goals

- Interviewer/hr./ Company owners or its representative:

  Make an asynchronous interview and put a number of questions for this interview

  Send invitations to each person applying for a job

  Receive a report on the personality traits and communication skills of each participant
- Interviewee:

  Receive an invitation for this interview

  Confirm entry by name and email

  Answer the questions via webcam and microphone on their mobile device or computer

  Confirm the answer to all questions

1.8 Related Work

There are a lot of video interview platform nowadays, but they do not have as much features as we apply in our system, the y mostly concerned by keeping all work in one place for the recruiter. The most popular are:
TestGorila: it is mostly concerned by keeping all work in one place for the recruiter, they facilitate the testing process by once the recruiter enter the job name, the website shows him a list of tests that related to this job role, the recruiter then choose a one or multiple tests and send them to the candidates.

EasyHire: Also it is concerned by keeping all work in one place for the recruiter, its AI system has just emotion detection part.

1.9 Project outline

We offered two literature reviews in chapter 2 and chapter 3, one for emotion recognition analysis techniques, and the other one for lie detection techniques. Chapter 4 includes all proposed techniques for interviewee feedback and analysis, and the results were discussed in chapter 6. And chapter 5 briefly explains the interaction of our websites. And chapter 7 shows the conclusion of the platform and the expected future work.

Chapter 2

Literature Review on Emotion Analysis

The Huge growth of online data at exponential level gives computer researchers numerous new challenges and opportunities. IVI's current model is to analyze interviewee's feelings and emotions during an interview. This is done by exploiting emotion recognition and analysis strategies. Emotion recognition is the computational study of people's feelings, attitudes and emotion toward something. Emotion recognition techniques are categorized into knowledge-based, statistical machine learning-based and hybrid techniques.

The structure of this chapter is as follows: **Section 1** describes the current emotion recognition analysis challenges for our platform. **Section 2** is the core of this literature review, and it covers the various techniques and approaches used in emotion recognition such as knowledge-based, statistical-based and hybrid techniques. Finally, **Section 3** shows the different emotion analysis application

2.1 Emotion Detection Challenges

The process of identifying human emotion is difficult for most people. People vary widely in their accuracy at recognizing the emotions of others, also there are different methods people use based on feelings and feelings are hard to formulate which make it tough for a computer to understand it.

The difficulties are compounded by several other challenges some of which are discussed below:

- ➔ **The Data Problem:** There is a lack of a benchmark datasets relevant to our platform for emotion recognition. Interview analysis relative to emotion recognition mostly fails due to the lack of interviewee emotion.

- ➔ **The Language Problem:** The majority of the work on emotion recognition has been done supporting English language. Most interviews and emotion datasets only support English where other widely used languages are not supported.

- ➔ **The Accent detection:** speech emotion detection may be affected by a strong accent which misleads the output as a different emotion.

- ➔ **Facial obstruction:** usually interviewees wear glasses with thick frames which impede eye feature extraction which causes a setback to the emotion recognition accuracy.

The use of technology to help people with emotion recognition is a relatively nascent research area. Decades of scientific research have been conducted developing and evaluating methods for automated emotion recognition. There is now a broad literature proposing and evaluating hundreds of different kinds of methods, leveraging techniques from multiple areas, such as signal processing, machine learning, computer vision and speech processing.

The accuracy of emotion recognition usually improves when it combines different emotion extraction techniques together to integrate a multimodal system based upon facial-based, voice-based, text-based, gesture-based and vital-based where these combinations are usually called hybrid systems. These techniques can be broadly classified into knowledge-based techniques, statistical methods and hybrid methods.

2.2 Emotion Detection Techniques/Approaches

Emotion recognition analysis is categorized into three categories. Knowledge-based techniques sometimes referred to as lexicon-based techniques, utilize domain knowledge and the semantic and syntactic characteristics of language in order to detect certain emotion types.

- **Knowledge-Based Techniques**

This technique depends on handling forms of text-based emotion recognition extracted from speech of the interview, one of the advantages of this approach is the accessibility and economy brought about by the large availability of such knowledge-based resources. A limitation of this technique on the other hand, is its inability to handle concept nuances and complex linguistic rules.

Knowledge-based techniques can be mainly classified into two categories: dictionary-based and corpus-based approaches. Dictionary- based approaches find opinion or emotion seed words in a dictionary and search for their synonyms and antonyms to expand the initial list of opinions or emotions. Corpus-based approaches on the other hand start with a seed list of opinion or emotion words, and expand the database by finding other words with context-specific characteristics in a large corpus. While corpus-based approaches take into account context, their performance still varies in different domains since a word in one domain can have a different orientation in another domain.

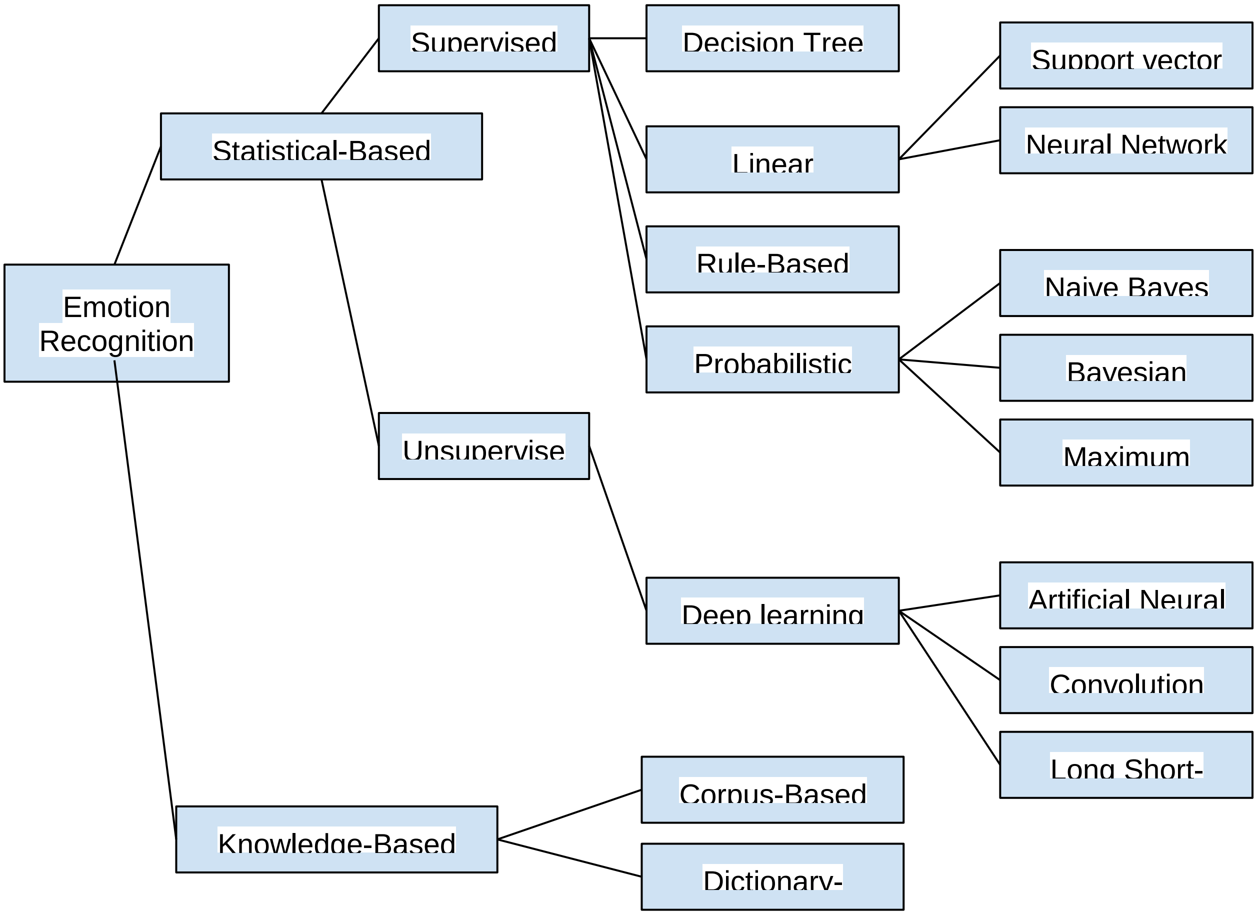
- **Statistical-Based Machine Learning Techniques**

Second technique is statistical methods (machine learning methods) which commonly involve the use of different supervised machine learning algorithms in which a large set of annotated data is fed into the algorithms for the system to learn and predict the appropriate emotion types.

Machine learning algorithms generally provide more reasonable classification accuracy compared to other approaches, but one of the challenges in achieving good results in the classification process is the need to have a sufficiently large training set.

Some of the most commonly used machine learning algorithms include support vector machines, neural networks under the category of linear classification of supervised learning algorithms and naive bayes, bayesian network and maximum entropy under the category of probabilistic classifiers of supervised learning algorithms.

Deep learning which is under the unsupervised learning algorithms is also widely employed in emotion recognition. Well known deep learning algorithms include different architectures of artificial neural networks such as convolutional neural networks, long short-term memory and extreme learning machines. The popularity of deep learning approaches in the domain of emotion recognition may be mainly attributed to its success in related applications such as in computer vision, speech recognition and natural language processing.



- **Hybrid-Based Techniques**

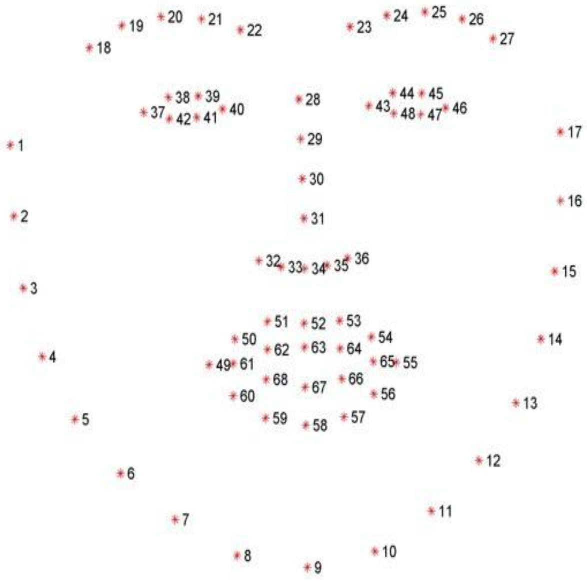
Third approach is hybrid approach, they are essentially a combination of knowledge-based techniques and statistical methods, which exploit complementary characteristics from both techniques. It has the potential to improve the emotion classification performance. The role of such knowledge-based resources in the implementation of hybrid approaches is better classification performance as opposed to employing knowledge-based or statistical methods independently. A downside of using hybrid techniques however, is the computational complexity during the classification process.

2.2.1 Facial Expression Based

Facial expressions play an important role in emotion recognition analysis and are used in the process of non-verbal communication, as well as to identify people. They are very important in daily emotional communication. They are also an indicator of feelings allowing a man to express an emotional state.

In the case of our platform, an interviewee’s face is the most exposed part of the body, allowing the use of computer vision systems to analyze the image of the face for recognizing emotion, light conditions and changes of head position are the main factors that affect the quality of emotions recognition systems.

Detecting facial landmarks is a subset of the emotion recognition problem, the position of each mark detects the location of the mouth, eyes and eyebrows relative to the emotion we are trying to predict.



Doing basic calculations like calculating the position of the eyebrows could result in a surprised emotion, the height of the lowermost, uppermost, leftmost and rightmost points of the mouth can calculate whether he is happy or neutral.

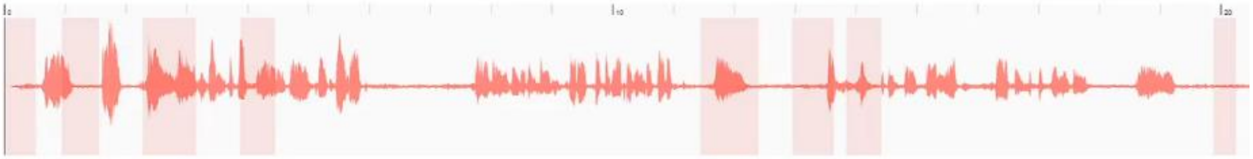
2.2.2 Voice Based

Emotional speech recognition involves automatically identifying the emotional or physical state of a human being from voice.

The importance of emotion recognition for human computer interaction is widely recognized, it is an important factor in human communication and it also provides feedback.

Voice emotion recognition focused on detecting the values of pitch, formant or cepstrum from the variation of speech according to changing emotion.

However we are more focused on detecting filler words and silence of the interviewee to detect the interviewee's tension during the interview and which support our emotion recognition system.



2.2.3 Vital Signals Based

Emotion Affects both human physiological and psychological status, in general emotion recognition methods could be classified into two major categories. One is using human physical signals such as facial expression, speech, gesture, etc, which has the advantage of easy collection and has been studied for years. However, the reliability can't be guaranteed, as it's relatively easy for people to control the physical signals like facial expression or speech to hide either real emotions especially during social communications.

For example, people might smile in a formal social occasion even if they are in a negative emotional state. The other category is using the internal signals - the physiological vital signals- which include the electroencephalogram, temperature, electrocardiogram, electromyogram, galvanic skin response, respiration, etc.

The nervous system is divided into two parts: the central and the peripheral nervous system. The peripheral nervous system consists of the autonomic and the somatic nervous systems. The autonomic somatic nervous system is composed of sensory and motor neurons, which operate between the central nervous system and various internal organs such as the heart, the lungs, the viscera and the glands.

All these systems change in a certain way when people face some specific situations. The physiological signals are in response to the central nervous system and the autonomic nervous system of the human body, in which emotion changes according to Connon’s theory (The James-lange theory of emotions: a critical examination and an alternative theory. By Walter B. Cannon, 1927).

One of the major benefits of the latter method is that the central nervous system and autonomic nervous systems are largely involuntarily activated and therefore cannot be easily controlled.

However, it was found that it was relatively difficult to precisely reflect emotional change by using a single physiological signal, therefore, emotion recognition using multiple physiological signals presents its significance in both research and real application.

However, using this method for our platform provides a challenge because hardware sensors to capture these signals are not available with the interviewee.

2.2.4 Gesture based

During conversations people are constantly changing nonverbal clues, communicated through body movements and facial expressions. The difference between the words people pronounce and our understanding of their content comes from nonverbal communication also commonly called body language.

The inner state of a person is expressed through elements such as iris extension, gaze direction, position of hands and legs, and the style of sitting, walking, standing or lying.

After the facial expression, hands are probably the richest source of body language information, for example based on the position of hands one is able to determine whether a person is honest (one will turn the hands inside towards the interlocutor) or insincere (hiding hands behind the back) Exercising open-handed gestures during conversation can give the impression of a more reliable person.

It's a trick often used in debates and political discussions. It is proven that people using open-handed gestures are perceived positively.

Head positioning also reveals a lot of information about emotional state. Researches indicates that people are prone to talk more if the listener encourages them by nodding. The pace of the nodding can signal patience or lack of it. In neutral position the head remains still in front of the interlocutor. If the chin is lifted it may mean that the person is displaying superiority or even arrogance. Exposing the neck might be a signal of submission.

The torso is probably the least communicative part of the body. However, its angle with the body is an indicative attitude. For example placing the torso frontally to the interlocutor can be considered as a display of aggression. By turning it at a slight angle one may be considered self-confident and devoid of aggression. Leaning forward especially when combined with nodding and smiling is the most distinct way to show curiosity.

The above considerations indicate that in order to correctly interpret body language as indicators of emotional state, various parts of body must be considered at the same time, however, this is not the case in our platform as our interviewee will be mainly recorded with his upper part of his body mainly facial expression will be the most representable way of capturing emotions features.

2.2.5 Hybrid Methods

The hybrid approach is a combination of both knowledge-based techniques and statistical-based techniques using one or more methods of the mentioned approaches (facial expression, voice based, vital signals based, gesture based) as features of emotion recognition analysis.

IVI’s combination techniques of facial expression and voice analysis in addition to lie detection (Chapter 3) provides valuable insight about our user, in IVI’s case, The interviewee, for further interview filtration process.

2.3 Emotion Analysis application

Emotion recognition is used in society for a variety of reasons.

Affective - provides artificial intelligence software that makes it more efficient to do tasks previously done manually by people, mainly to gather facial expression and vocal expression information related to specific contexts where viewers have consented to share this information.

For example, instead of filling out a lengthy survey about how you feel at each point watching an educational video, you can consent to have a camera watch your face and listen to what you say, and note during which parts of the experience you show expressions such as boredom, interest, confusion or smiling.

Other uses by affective include helping children with autism, helping people who are blind to reach facial expressions, helping robots interact more intelligently with people and monitoring signs of attention while driving in an effort to enhance driver safety.

Snapchat - filed a patent in 2015 describes a method of extracting data about crowds at public events by performing algorithmic emotion recognition on user’s geotagged selfies.

Emotient - was a startup company which applied emotion recognition to reading frowns, smiles and other expressions on faces to predict attitudes and actions based on facial expressions. Apple brought emotient in 2016 and uses emotion recognition technology to enhance the emotional intelligence of its products.

nViso - provides real-time emotion recognition for web and mobile applications through a real-time API. Visage technologies AB offers emotion estimation as a part of their Visage SDK for marketing and scientific research and similar purposes.

Eyeris - is an emotion recognition company that works with embedded system manufacturers including car makers and social robotic companies on integrating its face analytics and emotion recognition software, as well as with video content creators to help them measure the perceived effectiveness of their short and long form video creative.

Many products also exist to aggregate information from emotions communicated online, including via ‘like’ button presses and via counts of positive and negative phrases in text and affect recognition is increasingly used in some kinds of games and virtual reality, both for educational purposes and to give players more natural control over their social avatars.

Chapter 3

Literature Review on Lie Detection

Lies are everywhere. We live in a society full of lies. The consequences of lying impact any environment negatively whether at work, home or with friends. An interviewee lying about his skills can have a negative cost on the environment the interviewers are trying to make.

This platform shows the current lie detection challenges and lie detection techniques/approaches using cognitive techniques and imposing these techniques.

The structure of this chapter is as follows: **Section 1** gives an overview of Lie detection challenges. **Section 2** discusses lie detection techniques/approaches using cognitive load techniques and imposing these techniques in order to achieve a high lie detection model for our platform. **Section 3** concludes the platform overall method to assess interviews and provides remarks and future work.

3.1 Lie Detection challenges

A lie detection method can be effective only if it is based on sound theory about how people respond when they lie or tell the truth. When a theory can adequately predict what difference will occur between lying and telling the truth, an effective method can be designed to detect these differences.

Police manuals typically promote a ‘concern-based lie detection approach’ to detect deception. This approach assumes that people are more concerned when they lie than when they tell the truth. Resulting in nervous displays such as crossing the legs, shifting in the chair or grooming behavior.

This approach is limited because there is no compelling theoretical explanation as to why suspects would necessarily be concerned and display nervous behaviors when they lie. Neither is it clear why they should necessarily be unconcerned and stay calm when telling the truth.

Theoretically there are four challenges we will examine must be met in order to allow us to judge whether an interviewee is lying.

Relevance the first challenge for research in lie detection training is to create a deception situation that is relevant to the interview subject.

Stakes is to verify the observable clues to lying after creating the lie situation. There are clues that distinguish lies from truths.

Training once we have relevant and reliable deception material, we must be able to drive these clues into conclusions.

Testing is to ensure that there are adequate pre and post measures of the training phase and making sure the results are correct.

3.2 Lie Detection Techniques/Approaches

Researches developed two theory-driven approaches to discriminate between truth-telling and lying. Cognitive Load Techniques and imposing these techniques.

This approach is based on the assumption that lying is sometimes more cognitively demanding than truth-telling. Several aspects of lying contribute to this increased mental load. First, formulating the lie itself may be cognitively taxing.

Second, people are typically less likely to take their credibility for granted when they lie than when they tell the truth. When lying, people will be more inclined to monitor and control their demeanour so that they will appear honest to the lie detector than when telling the truth, which is cognitively demanding.

Third, because people do not take credibility for granted when they lie, they may monitor the interviewer's reaction more carefully in order to assess whether they are getting away with their lie.

Fourth, when people lie, they may be preoccupied by the task of reminding themselves to act and role-play, which require extra cognitive effort.

Fifth, people have to suppress the truth when they are lying and this is also cognitively demanding.

Finally, where activating the truth often happens automatically activating a lie is more intentional and deliberate, thus requires mental effort.

The above-mentioned reasons as to why lying is more cognitively demanding could give us insight into when it is more cognitively demanding. Lying is more cognitively demanding to the degree that these six principles are in effect.

For example, lying is likely to be more demanding than truth-telling only when the interviewees are motivated to be believed. Only under those circumstances can it be assumed that people take their credibility less for granted when lying than when telling the truth, and hence will be more inclined to monitor their own behaviour and/or the interviewer's reactions when lying.

Second, for lying to be more cognitively demanding than truth-telling, people must be able to recall their truthful activity easily when they lie and have a clear memory of it. Only when their knowledge of the truth is easily and clearly accessed will it be difficult for them to suppress the truth.

On the other side of the equation, truth tellers also need to have easy access to the truthful event for the task to be relatively undemanding. If people have to think hard to remember the target event when they tell the truth (e.g. because it was not distinctive or it occurred long ago and was either not meaningful or not rehearsed), their cognitive demands may exceed the cognitive demands that are required for fabricating a story.

3.2.1 Cognitive Load Techniques

It is assumed that the mere act of lying generates observable signs of cognitive load, some researchers argue that increased cognitive load will result in slower response times and recommend examining response times.

Others suggest that increased cognitive load results in a decrease in movements recommend observing hand, food, leg and eye blinking. This is the traditional cognitive lie detection approach and has its roots in the seminal paper written by Zuckerman, DePaulo and Rosenthal (1981).

3.2.2 Imposing Cognitive Load Techniques

This approach goes one step further, in this innovative approach additional cognitive demand is imposed on interviewees to enlarge the observable cognitive differences between lying and truth-telling. The core of this approach is that lie detectors could exploit the increase in cognitive load that people experience when they lie by introducing mentally taxing interventions.

People require more cognitive resources when they lie than when they tell the truth to produce their statements, and therefore will have fewer cognitive resources left over to address these mentally taxing interventions when they lie than when they tell the truth.

This should result in more pronounced differences between lying and truth-telling in terms of displaying signs of cognitive load—e.g. more stutters and pauses, slower speech, slower response times, less quality details, inconsistencies, fewer movements—when these cognitively demanding interventions are introduced than when such interventions are not introduced.

3.3 Summary and Concluding Remarks

There are numerous research efforts about each topic we discussed, on the other hand there are very few researches conducted under the relying domain of interviews. Automating the interview process could be a huge saving of time and cost for major companies.

IVI's proposed method is a combination between emotion analysis and lie detection to provide a feedback of the interviewee.

Chapter 4

The Proposed Platform Approaches

In this chapter we will discuss the platform approaches in analyzing emotion from facial expression and voice analysis to support the emotion status and provide more feedback for lie detection and tension analysis during the interview.

The structure of this chapter is as follows: **Section 1** discussing the methods of facial expression emotion analysis, preprocessing modules and feature extraction and the architecture of the proposed methods. **Section 2** overviewing the process of disfluency voice analysis and segmentation to detect tension. **Section 3** explains lie detection approach using cognitive load analysis and discussing the thresholds. **Section 4** explaining skills extraction from speech. **Section 5** overviewing the overall platform statistical review for the interviewee.

4.1 Facial Expression Emotion Analysis

Emotion recognition is a technique used in software that allows a program to "read" the emotions on a human face using advanced image processing. Companies have been experimenting with combining sophisticated algorithms with image processing techniques that have emerged in the past ten years to understand more about what an image or a video of a person's face tells us about how he/she is feeling and not just that but also showing the probabilities of mixed emotions a face could have.

The model analysis is based on mini-Xception network architecture that is pre-trained on the image Net dataset and included in keras library since it gave us the best accuracy.

The proposed architecture categorizes each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

The human accuracy for classifying an image of a face in one of 7 different emotion is approximated around 65% where our platform needs to be more robust than that accuracy and computationally efficient, the state-of-the-art methods in image-related tasks such as image classification and object detection are all based on convolutional neural networks. These tasks require CNN architectures with millions of parameters.

In this platform we propose an implementation of a general convolutional neural network framework for designing real-time Emotion recognition and facial expression system.

4.1.1 Preprocessing

Preprocessing images for our platform consider the processes of transforming an input data in our case image to produce an output that is normalized for unifying all images under the same domain.

The platform takes every frame of the imputed interview’s video and resizes it’s width equal to 300 using imutils modules. Then convert it from BGR to Grayscale using OpenCv library.

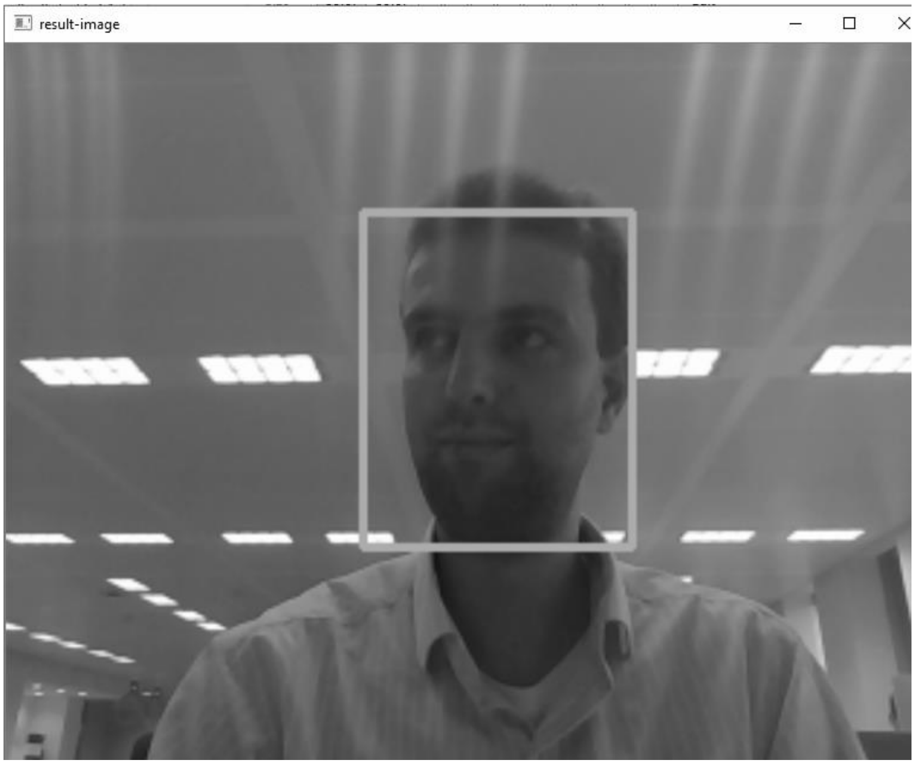
And the output is then passed through a Haarcascade face detection model with the following parameters:

- scaleFactor = 1.1

- minNeighbors = 5
- minSize = (30, 30)

The previous parameters were handpicked by trial and error and they proved to work best on webcams, which supports our main goal of detecting the face of the candidate via a webcam where he is positioned in front of the camera with a close distance and in an environment with clear lighting.

The output looks something like this:



The detected face is then cropped and resized to a 64x64 image.

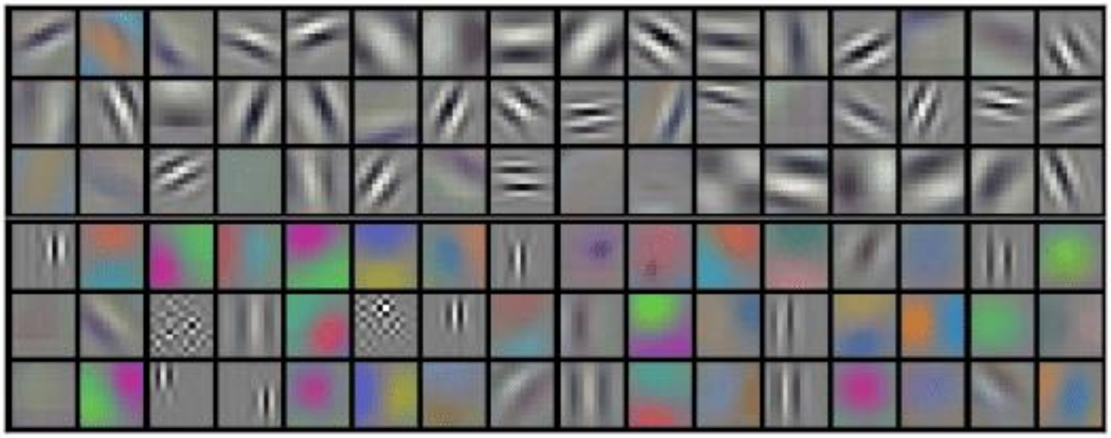
We then used Keras's img_to_array() function which turns an image into an array.

And then the output is finally passed to our Emotion Detection Model.

4.1.2 Feature Extractions

The characteristic of a convolutional neural network is to alter the input in a way to reduce the parameters in addition to transforming the input in a way to maximize the changes so called features which define the input.

This is done by altering the kernel values and the pooling layers.

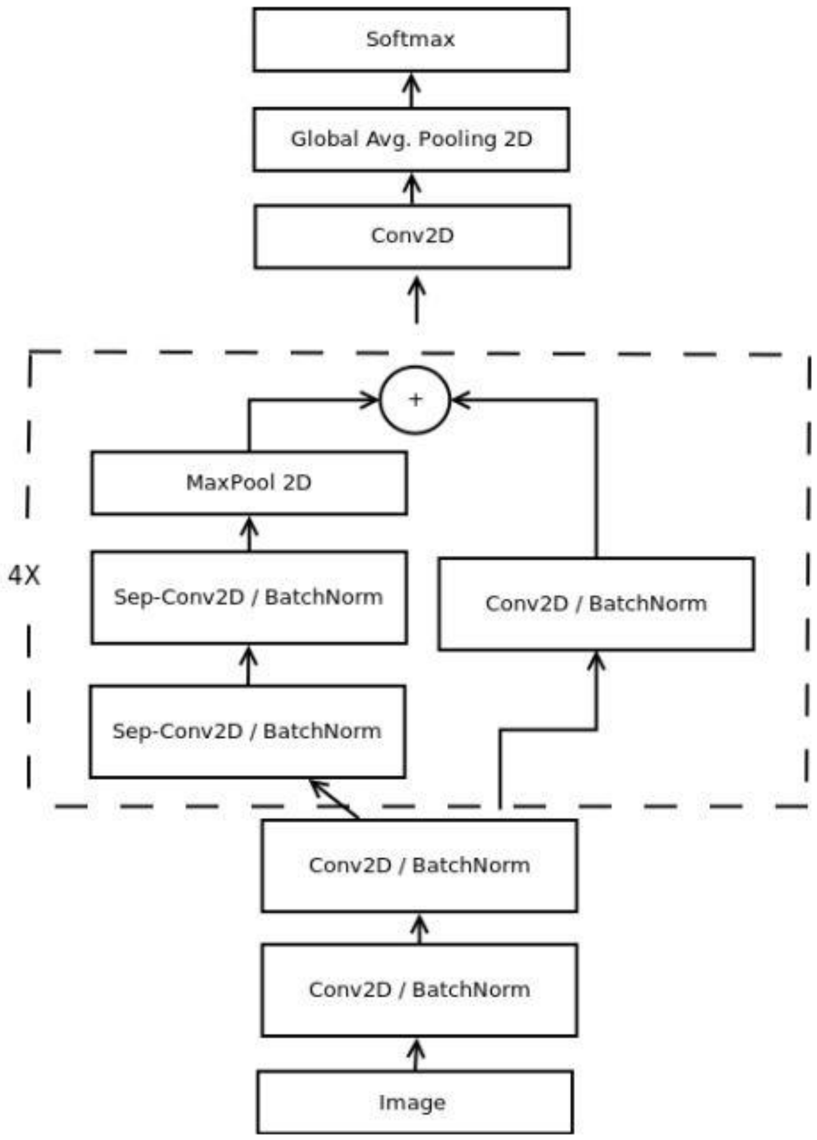


This technique allows the architecture to find edges in the preprocessed image of our platform which corresponds to facial features such as mouth, eyes and nose.

4.1.3 Architecture

The architecture relies on the idea of eliminating completely the fully connected layers, this architecture schema was trained with the ADAM optimizer.

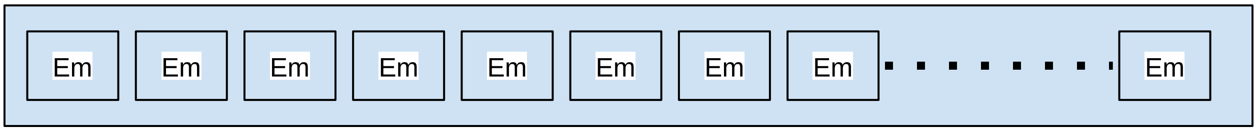
Eliminating fully connected layers was achieved by having the last convolutional layer the same number of feature maps as the number of classes and applying a softmax activation function.



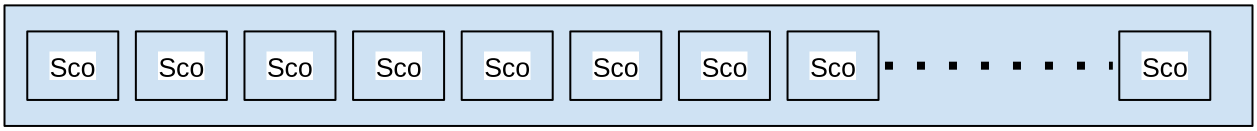
4.1.4 Output

The output of this model is the probability of each emotion detected.

In normal case scenarios we would use an argmax function so that the output is the emotion that our model is most confident in. but in our case we chose to set a threshold if the confidence of any of the emotions passes that threshold it is then added to an array and that array is added to an array of list of emotion for each frame of the entire video



When the process is finished for all of the frames, we pass through the list of frames emotions and add a numeric score for each frame based on the detected emotions in that frame



The method that we have chosen to add a score to each frame is pretty simple and straightforward.

Each frame starts with a score of zero, and the following snippet of pseudo-code shows how it is done step by step.

```
Score = 0
If Frame has Anger:
     Score -= 1
     If Frame has Neutral:
          Score += 0.5
If Frame has Happy:
     Score += 2
     If Frame has Neutral:
          Score -= 1
If Frame has Sad:
     Score -=1
     If Frame has Neutral:
          Score += 0.5
```

And then this scoring method is repeated for each frame until the list of frame scores for the entire video is finished. It is then normalized so that the scores are between the intervals of [-1, 1] instead of [-2.2]. -1 being a very negative emotional state, and 1 being a very positive emotional statues

And then we calculated the overall score based on the assumption that too much smiling is bad and too serious is bad as well. The optimal score comes from the average of distances between all the scores and the 0.5 region that represents [neutral, happy]

```
emo_stat_arr = [x / 2 for x in emo_stat_arr]
emo_stat_arr = [x - 0.5 for x in emo_stat_arr]
for i in range(len(emo_stat_arr)):
      if emo_stat_arr[i] < 0:
                 emo_stat_arr[i] = -emo_stat_arr[i]
emo_stat_arr = [x / 1.5 for x in emo_stat_arr]
emo_stat_arr_avg = statistics.mean(self.emo_stat_arr)
emo_score = 1-emo_stat_arr_avg
```

4.2 Disfluency Voice Analysis

Conversation is governed by expectations of timely responding. Violations of these expectations are grounds for inference by other participants. These inferences may be at odds with identities respondents try to project. In job interviews, candidates' responses are used to make hiring decisions. Candidates trade off between (1) delaying response initiation to search for an appropriate response at the risk of appearing inept and (2) responding quicker but less appropriately. In a corpus of job interviews, response delays predicted the probability of inappropriate initial responses and decreased hireability ratings, illustrating how unintended aspects of conversational delivery can entail social and institutional consequences beyond the conversation itself.

In our disfluency detection model we mainly focused in detecting the Aaahs and Umms of the candidate during answering the questions, it comes without saying that a large amount of disfluencies in a small time period means that the candidate is unsure about his answer. Which plays a major in detecting whether that answer is correct at all or not.

4.2.1 Voice segmentation module

Our model takes an Audio Wav File version of the video.

We used a code that we have borrowed from liberosa library examples to clean the audio, and by cleaning the audio we mean separating the foreground noises (the voice of the actual speaking of the candidate) and the background noises (Car sounds, Air Conditioner noise, background murmuring...etc).

We then extract the features from the foreground noise file also using the liberosa library, the following code snippet function shows the features that are extracted and how it was extracted.

```
def feat_ext(file_name,num_mfcc=40):
        X, sample_rate = librosa.load(file_name)
                mfccs = librosa.feature.mfcc(S=librosa.power_to_db(mel),n_mfcc=num_mfcc)
        return mfccs,X, sample_rate
```

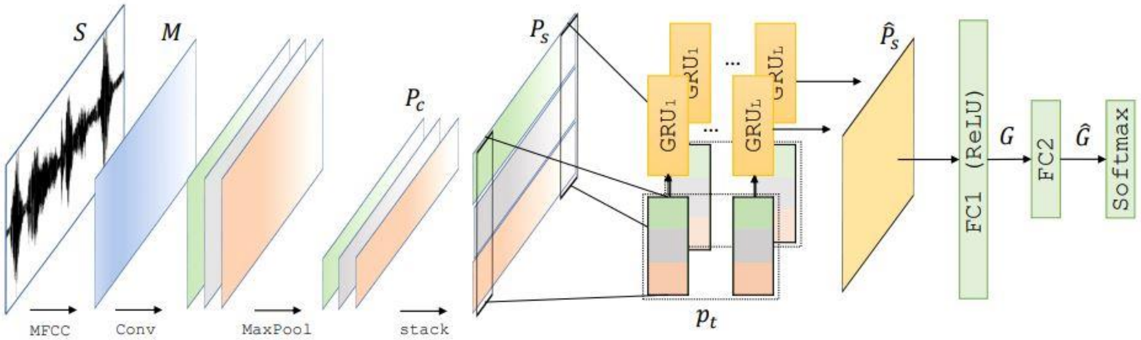
And then we pass the mfccs/feats variable to a function that takes in the features of the entire audio segment and returns a list of features in size (128, 201), length of padding, number of contiguous segments and [(start,end)] for all the random windows.

We then pass those features through our Umm Segmentation model and return the time segment of all the Umms and Aahs that the candidate has said.

4.2.2 The Model Architecture

For this task we have used a trained CRNN model that was inspired by the paper titled INCREASE APPARENT PUBLIC SPEAKING FLUENCY BY SPEECH AUGMENTATION submitted in August 2019

The model was trained on a curpos called SWITCHBOARD which is a corpus of spontaneous conversations which addresses the growing need for large multi-speaker databases of telephone bandwidth speech. The corpus contains 2430 conversations averaging 6 minutes in length; in other words, over 240 hours of recorded speech, and about 3 million words of text, spoken by over 500 speakers of both sexes from every major dialect of American English.



4.2.3 Disfluency Score

When we calculate the anxiety score we put main things into consideration.

1- Number of disfluencies:

Based on TedX public speech statistics the average person says filler words once every 12 seconds. Thus we took this as our standard.

If the person says avg of 0 filler word he gets a perfect score of 1.
If they say avg of 1/12 per second he gets a score of 0.5.
If they say avg of ⅙ or more per second then he gets a score of 0
And everything in between is just calculated based on the linear function (Score = 6 * Disfluencies Per Second)

2- Length of the disfluencies.

Based on the assumption that 1 second of disfluency is long enough to count as worst for an interview; we give all the disfluencies that lasts 1 second or longer the worst score.
And disfluencies that are barely audible and last 0.2 second the best score.
Then we take the average of all the scores and count that as the Score of disfluencies based on how long each of them lasted.

```
ummsPerSecond = umms_count/timeInSeconds


if ummsPerSecond >= 2/12:
      disfc_score = 1
Else:
      disfluency_count_score = 6*ummsPerSecond


disfluency_count_score= 1 - disfluency_count_score


durations = []
for i in range(umms_count):
      if end_time[i] - start_time[i] > 1:
            durations.append(1)
      elif end_time[i] - start_time[i] < 0.2:
            durations.append(0)
      else:
            durations.append(end_time[i] - start_time[i])


 duration_score = statistics.mean(durations)


 duration_score = 1 - duration_score


 filler_score = (duration_score + disfluency_count_score) / 2
```

4.3 Anxiety Detection

The fact is, people lie. These can be small, irrelevant lies, and they can also cover some of the bigger and more significant things. It is also a fact that a lie detector has not yet been designed to operate with great precision.

The main reason is that the human body behaves differently in stressful situations. On the other hand, experienced and good liars have an exceptional ability to control reactions and it is difficult and even impossible to determine whether or not a statement is true.

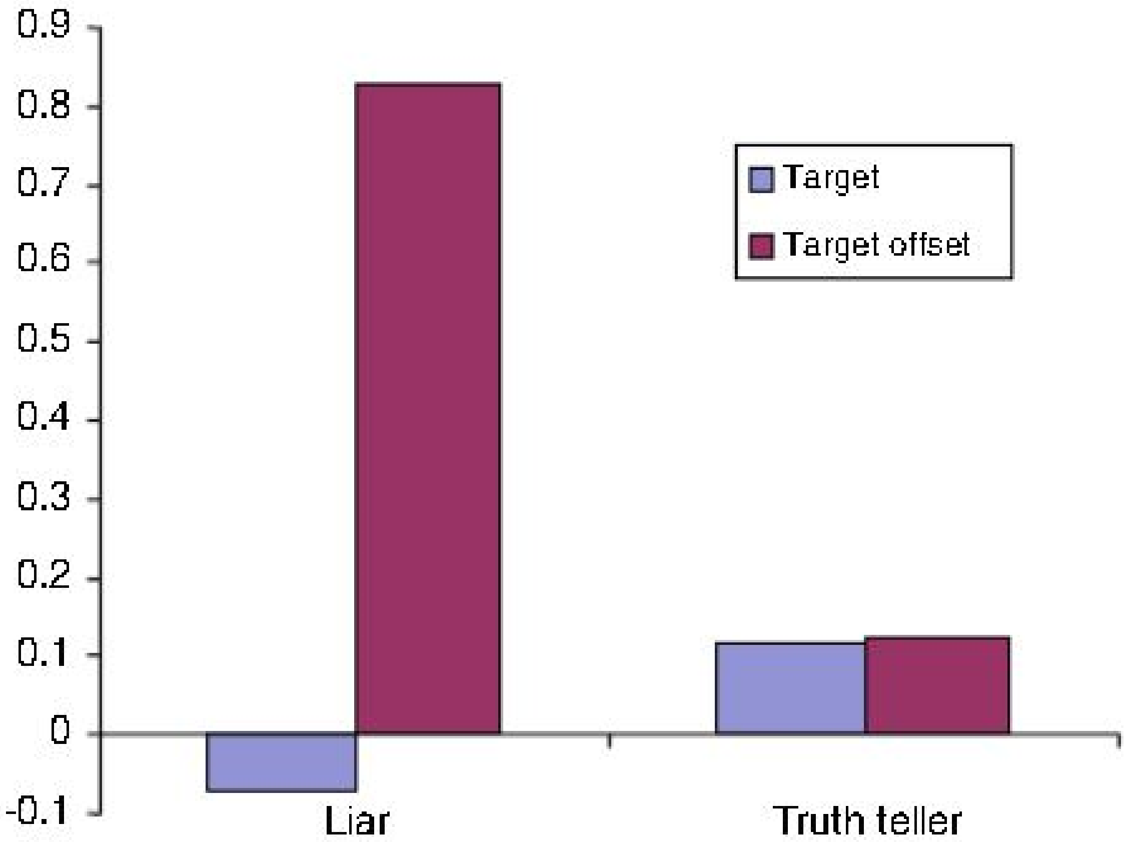
It is necessary to detect the truth of a person's testimony based on the facial expressions a person makes when answering the questions asked. Blinking, and lip squeezing are the parameters extracted and processed.

- **Anxiety derived from blinking**

Based on a paper published in springer by the title “Blinking During and After Lying” in January 2008, when liars experience cognitive demand, their lies would be associated with a decrease in eye blinks, directly followed by an increase in eye blinks when the demand has ceased after the lie is told. A total of 13 liars and 13 truth tellers lied or told the truth in a target period; liars and truth tellers both told the truth in two baseline periods. Their eye blinks during the target and baseline periods and directly after the target period (target offset period) were recorded. The predicted pattern (compared to the baseline periods, a decrease in eye blinks during the target period and an increase in eye blinks during the target offset period) was found in liars and was strikingly different from the pattern obtained in truth tellers. They showed an increase in eye blinks during the target period compared to the baseline periods, whereas their pattern of eye blinks in the target offset period did not differ from baseline periods.

And this is exactly what we did for detecting abnormal frequencies of eye blinking which may lead to a possible chance of a lie in the candidates interview video.

And we will follow the same again but also when it comes to lip squeezing later



Changes in blink rate per second during the target period and directly after the target period (target offset)

4.3.1 Blink Detection Algorithm

To build our blink detector, we'll be computing a metric called the *eye aspect ratio* (EAR), introduced by Soukupová and Čech in their 2016 paper, *Real-Time Eye Blink Detection Using Facial Landmarks*.

Unlike traditional image processing methods for computing blinks which typically involve some combination of:

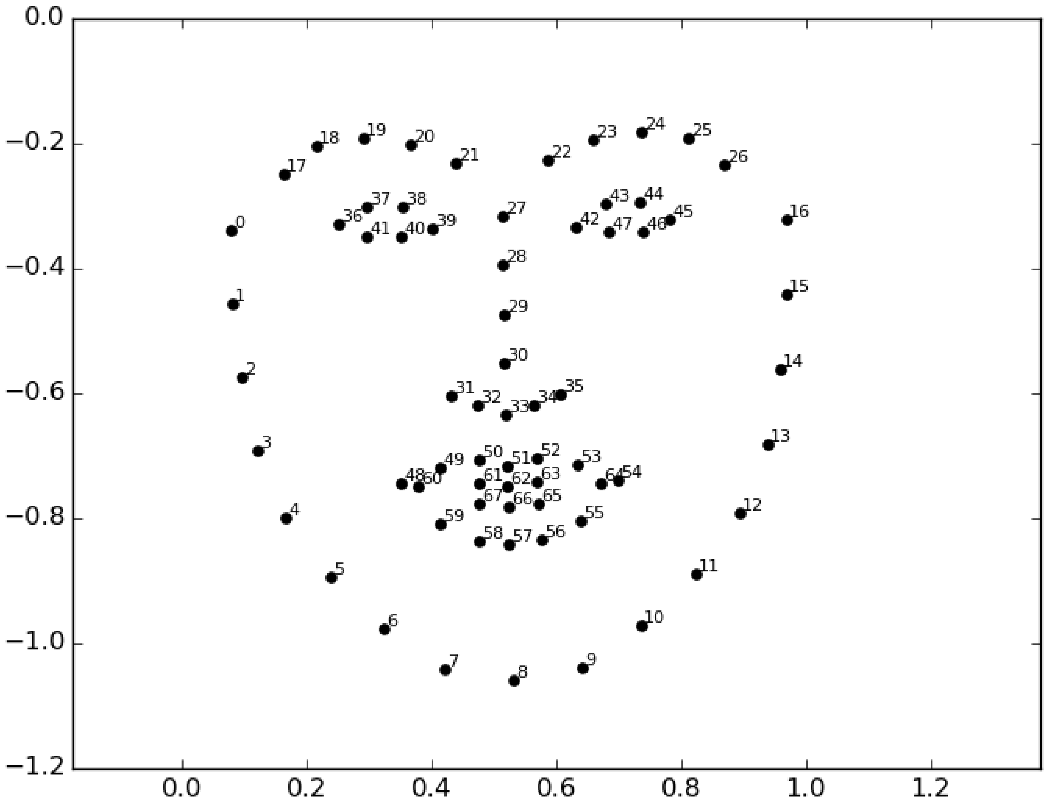
1. Eye localization.
2. Thresholding to find the whites of the eyes.
3. Determining if the "white" region of the eyes disappears for a period of time (indicating a blink).

The eye aspect ratio is instead a *much more elegant solution* that involves a *very simple calculation* based on the ratio of distances between facial landmarks of the eyes.

This method for eye blink detection is fast, efficient, and easy to implement.

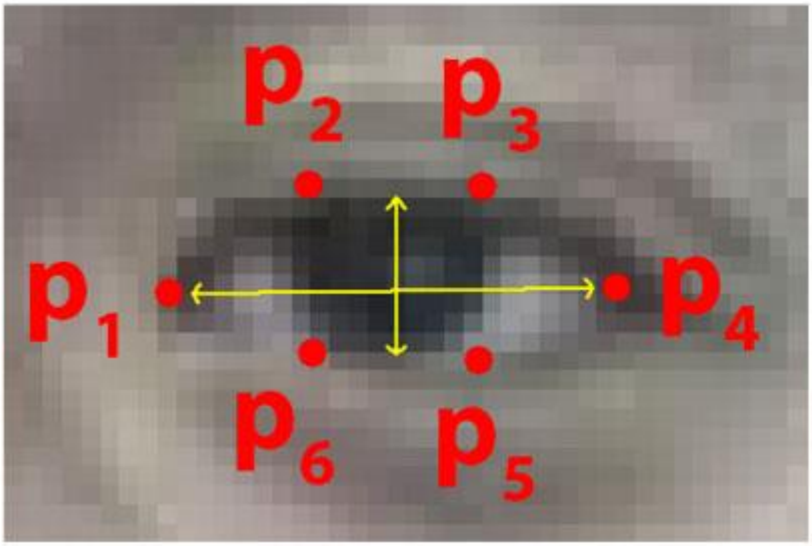
Using the dlib library we can apply facial landmark detection to localize important regions of the face, including eyes, eyebrows, nose, ears, and mouth

This also implies that we can extract specific facial structures by knowing the indexes of the particular face parts:



In terms of blink detection, we are only interested in two sets of facial structures — the eyes.

Each eye is represented by 6 (x, y)-coordinates, starting at the left-corner of the eye (as if you were looking at the person), and then working clockwise around the remainder of the region:



Based on this image, we should take away on key point:

There is a relation between the width and the height of these coordinates.

Based on the work by Soukupová and Čech in their 2016 paper, Real-Time Eye Blink Detection using Facial Landmarks, we can then derive an equation that reflects this relation called the eye aspect ratio (EAR):

$$\mathrm{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

Where p1, …, p6 are 2D facial landmark locations.

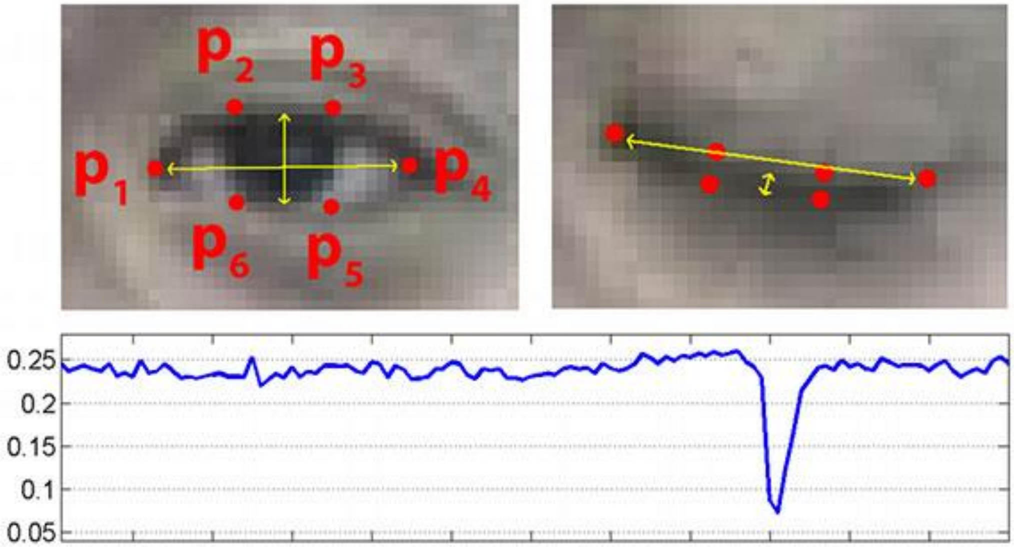
The numerator of this equation computes the distance between the vertical eye landmarks while the denominator computes the distance between horizontal eye landmarks, weighting the denominator appropriately since there is only one set of horizontal points but two sets of vertical points.

Why is this equation so interesting?

Well, as we'll find out, the eye aspect ratio is approximately constant while the eye is open, but will rapidly fall to zero when a blink is taking place.

Using this simple equation, we can avoid image processing techniques and simply rely on the ratio of eye landmark distances to determine if a person is blinking.

To make this clearer, consider the following figure from Soukupová and Čech:



On the top-left we have an eye that is fully open — the eye aspect ratio here would be large(r) and relatively constant over time.

However, once the person blinks (top-right) the eye aspect ratio decreases dramatically, approaching zero.

The bottom figure plots a graph of the eye aspect ratio over time for a video clip. As we can see, the eye aspect ratio is constant, then rapidly drops close to zero, then increases again, indicating a single blink has taken place.

And thus we have finally detected a blink from a video.

Blind Frequencies:

After detecting the blink itself we can easily detect the timestamp of where it happened and build an array of the blinks timestamps. Then can use the distance between blinks to calculate the frequencies and detect any abnormalities in the blinking pattern. Using the methods mentioned above in the **Abnormal Blinking & Lip pressing Model** section of this documentation.

4.3.2 Lies derived from Lip Squeezing

Why we detect Lip Squeezing:

Based on a blog posted in “Psychology Today” with the title “Lips Don’t Lie”

Lips convey a lot of information that is often ignored or not even observed. Rich with nerves and highly vascular, the lips react in real time to the world around us. So when people receive bad news or witness a horrific event their lips begin to disappear, becoming very thin as vasoconstriction takes place. Under extreme stress they disappear completely or are compressed together.

In relationships, couples will immediately notice when their partner has issues because they notice the tightening or compressing of the lips. Even kisses will seem different under stress as blood flow is restricted which affects their fullness, warmth, and pliability. Our lips react to the reality of the moment and communicate accurately our feelings and sentiments to others.

Because disappearing or compressed lips are universal behaviors, controlled by the limbic system, these are behaviors that can be relied upon and are authentic. We don’t realize how our lips look and feel, but others will notice.

Lip biting, like lip compression, is one of the ways that we pacify ourselves when we are stressed. It helps to relieve tension that may be minor and transitory. However, as I note in my book "Clues to Deceit," when something more significant is bothering us, our limbic system, in reaction to events, will compel the lips to narrow and disappear if the stress is significant enough.

We can use these behaviors to assess the level of comfort and discomfort noticeable on those we are observing. Students getting ready to take tests will demonstrate their stress level with these behaviors as will individuals who are suddenly confronted with disagreeable circumstances. As an FBI special agent, I used these behaviors (lip compression, disappearing lips) to determine what specific subjects stressed the interviewee suggesting there might be “guilty knowledge."

For example, when I would ask someone, “Do you own a gun?” They would say “yes,” and then I would notice the lips would disappear or be compressed slightly. Then I would ask, “Do you own a “Smith and Wesson revolver?” their lips would not react all that much. However, when I asked, “Do you own a “Glock pistol?” knowing that this was the weapon found at the crime scene and unknown to the public, I noticed that the lips became really narrow and compressed and that the corners of the mouth would also turn downward .

This to me was extremely significant in verifying that this individual was severely stressed by the question and most likely by his guilty knowledge.

In summary, Lip squeezing before or after answering the question means deceit in the answer.

Lip Squeezing Detection:

Now that we have talked about WHY we need to detect lip pressing, we will talk about HOW to detect the lip pressing.

We will basically follow the same steps we have done to detect the blinks, but this time with lips. We use the lip aspect ratio and do calculations based on the ratio of distance between facial landmarks of the lip using the dlib library.

This time we should take two additional things into consideration, the time period that the lip was squeezed for, because of course when the candidate talks he tends to close his lips and we have to avoid that by setting a threshold for the period of time that the lips are squeezed together. The second thing we should take into consideration is that we should make sure that the smiles are not counted as lip squeezes. Because interviewees tend to smile during the interview to show confidence and to give positive vibes during the interview.

So we detect lip squeezing before and after the question answering and add that part to the report to notify the employee of abnormal lips movement during the interview.

4.3.3 Hybrid Techniques

We will merge the Abnormal Blinks Frequencies and Lip Squeezing detections, we will use the output of both models and if both return abnormalities then we will add to the report that there is a Lie detected in this video.

4.3.4 Scoring

A score will be added based on if the model detects blink abnormalities only, lip squeezing abnormalities only, both, or none at all. And that also will be added to the overall score of the question.

To calculate the overall score based on anxiety we had first calculate the anxiety score based on blinking and anxiety score based on lips pursing and then averaging them.

We first cut the video into 4 quarters, and we calculated the ratio blinking count between the last quarter and each of the first 3 quarters. And if that ratio is more than 0.2 then we increase the anxiety score by 0.33.

And to calculate the anxiety score based on lips pursing we used the same strategy of cutting the video into 4 quarters and calculating the lip count in the first and last quarters. If the count of lip pursing in the first quarter is more than the count of each the second and third quarters we add an anxiety score of 0.25. And repeat the process for the last quarter.

This next code snippet can make the process more understandable:

```
def chunkIt(self, seq, num = 4):
        avg = len(seq) / float(num)
        out = []
        last = 0.0

        while last < len(seq):
            out.append(seq[int(last):int(last + avg)])
            last += avg

        return out
```

```
def detect_if_anx(self,history_of_blinks, history_of_lips):
        quad_blink_hist = self.chunkIt(history_of_blinks,4)
        quad_lip_hist = self.chunkIt(history_of_lips,4)
        blink_anx = 0
        lip_anx = 0
        quad_blink_hist_sum = []
        quad_blink_hist_sum.append(sum(quad_blink_hist[0]))
        quad_blink_hist_sum.append(sum(quad_blink_hist[1]))
        quad_blink_hist_sum.append(sum(quad_blink_hist[2]))
        quad_blink_hist_sum.append(sum(quad_blink_hist[3]))

        if quad_blink_hist_sum[3] > (quad_blink_hist_sum[0] +
quad_blink_hist_sum[0] * 0.25):
            blink_anx+= 0.33
        if quad_blink_hist_sum[3] > (quad_blink_hist_sum[1] +
quad_blink_hist_sum[1] * 0.25):
            blink_anx+= 0.33
        if quad_blink_hist_sum[3] > (quad_blink_hist_sum[2] +
quad_blink_hist_sum[2] * 0.25):
            blink_anx+= 0.33

        quad_lip_hist_sum = []
        quad_lip_hist_sum.append(sum(quad_lip_hist[0]))
        quad_lip_hist_sum.append(sum(quad_lip_hist[1]))
        quad_lip_hist_sum.append(sum(quad_lip_hist[2]))
        quad_lip_hist_sum.append(sum(quad_lip_hist[3]))
```

4.4 Skills Extraction

The reason an interviewer cares about a candidate’s skills and experience is if they believe such skills and experience will make them good at the job they are trying to fill. For them, everything is about your potential, and as such you should emphasize this as much as possible.

And since focusing on the mentioned skills during the interview can play an important role on accepting or rejected the candidate, extracting those skills from his speech and highlighting them in the summary will help the employer to quickly grasp the strength points of the candidate without having to watch the entire video or read through the entire Speech to Text version of the interview.

4.4.1 Speech to Text API

Speech recognition (also known as voice recognition) is the process of converting spoken words into computer text. The user speaks into a microphone and the computer creates a text file of the words they have spoken.

We used a simple `SpeechRecognition` library existing in python to make a text version of the candidate's answer.
And by only using very few lines of code we had an accurate reliable speech to text program ready to serve our project.

This text version is then used to detect skills mentioned in the interview and highlighting them for the employee.

4.4.2 Extracting Skills from Unstructured Text

Simply having a database of skills and passing through the text to check every word if it is in the database or not is the first thing that comes into mind when trying to extract skill from unstructured (speech) text. But of course this won't work for many reasons, skills having different abbreviations, or said in different contexts are the two main reasons.

So the optimal solution here was trying to detect skills based on their context in the sentence.

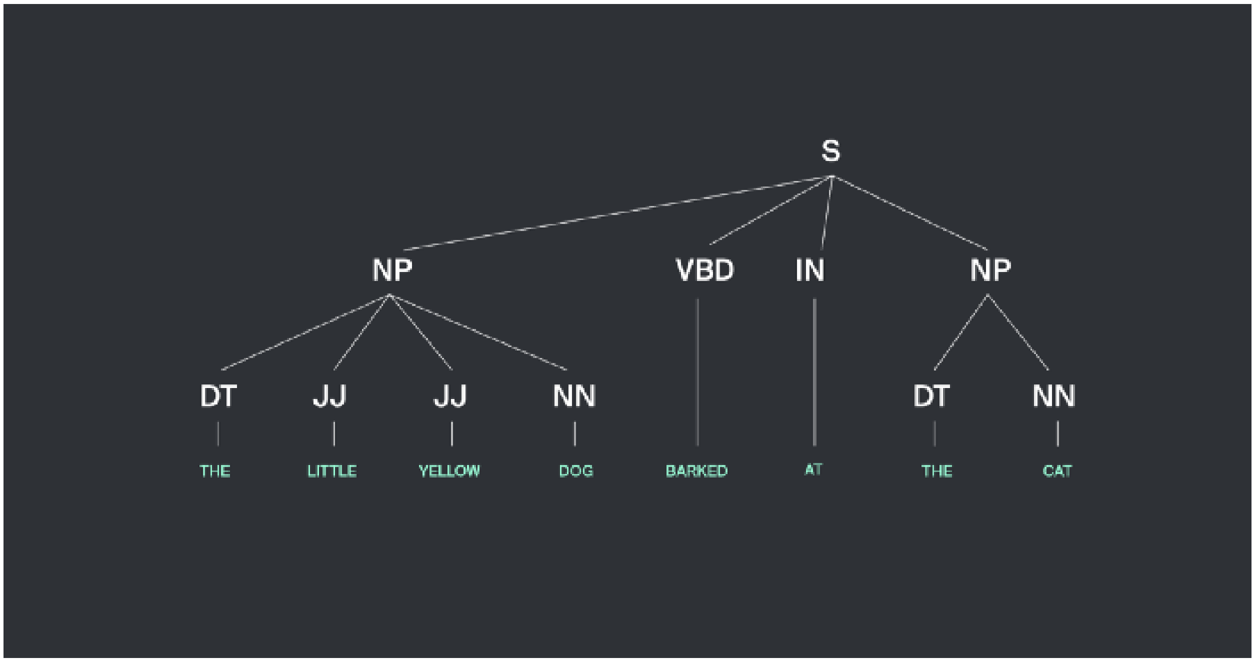
Example:

CV: Data scientist, hands on expertise in machine learning, big data, development, statistics and analytics. My team of data scientists implemented Python machine learning model ensembles, stacking, and feature engineering demonstrating high accuracy rates in predictive analytics. Created a recommender system using Doc2Vec words embeddings and neural networks.

Extracted professional skills: machine learning, big data, development, statistics, analytics, Python machine learning model ensembles, stacking, feature engineering, predictive analytics, Doc2Vec, words embeddings, neural networks.

Step 1: Parts of speech tagging

The task of entities extraction is a part of text mining class problems — extracting some structured information from an unstructured text. Let us take a close look at the suggested entities extraction methodology. As far as skills are mainly present in so-called noun phrases the first step in our extraction process would be entity recognition performed by NLTK library built-in methods (checkout Extracting Information from Text, NLTK book, part 7). Part of speech tagging method extracts noun phrases (NP) and builds trees representing relationships between noun phrases and the other parts of the sentence. NLTK library has a number of tools performing such phrase decomposition.

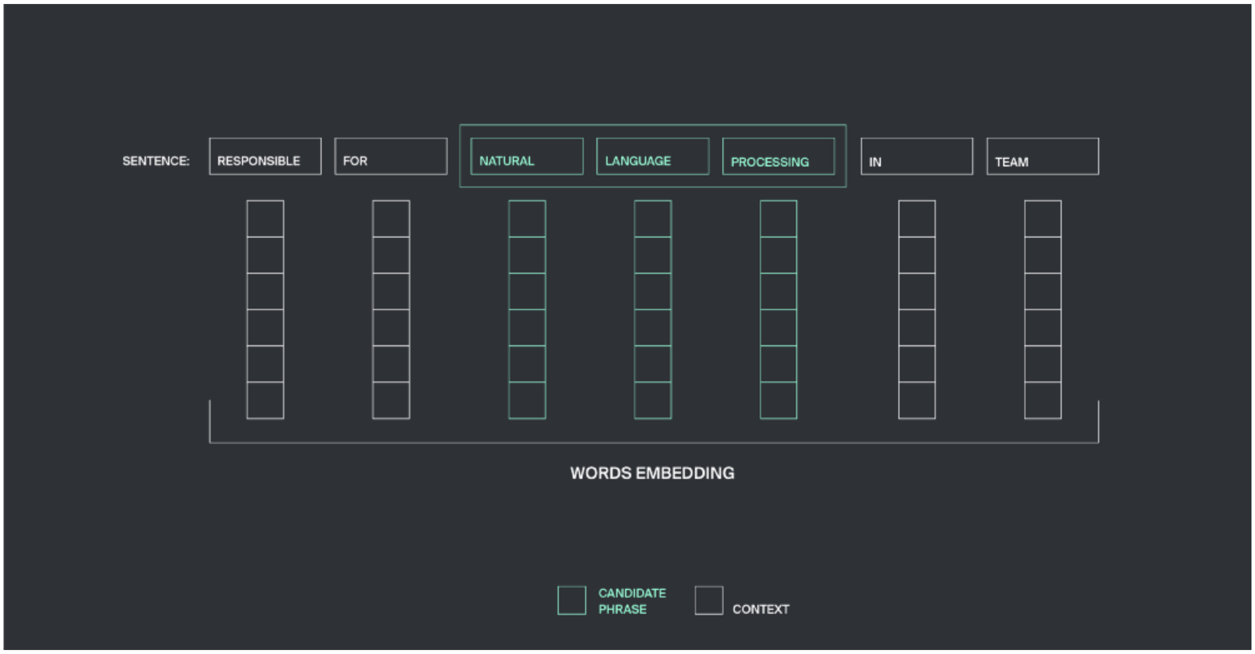


We can define a model as a regular expression giving the sentence decomposition (for example, we can define a phrase as a number of adjectives plus a noun) or we can teach a model on a labeled number of texts from NLTK with extracted noun phrases examples in them. This step results in receiving a number of entities among which some are the target skills and some are not — besides skills CV could contain some other entities such as places, persons, objects, organizations, whatever.

Step 2: Deep learning architecture for candidates' classification

The next step is entities classification. Here the objective is quite simple — to tell skills from “not skills”. The set of features used for training is composed regarding the structure of the candidate phrase and the context. Obviously, to train a model we had to create a labeled training set. Popular part of speech taggers (NLTK POS tagger, Stanford POS tagger) often make mistakes in the CV’s phrases tagging task. The reason is that often a CV text neglects grammar in order to highlight experience and to give it some structure (people start sentences with a predicate, not with a subject, sometimes phrases miss appropriate grammatical structure), a lot of words are specific terms or names. We had to write our own POS tagger solving the aforementioned problems.

The classification is performed with a Keras neural network with three input layers each designed to take a special class of data. The first input layer takes a variable length vector comprised of the described above features of the candidate phrases which could have an arbitrary number of words. This feature vector is processed with an LSTM layer.

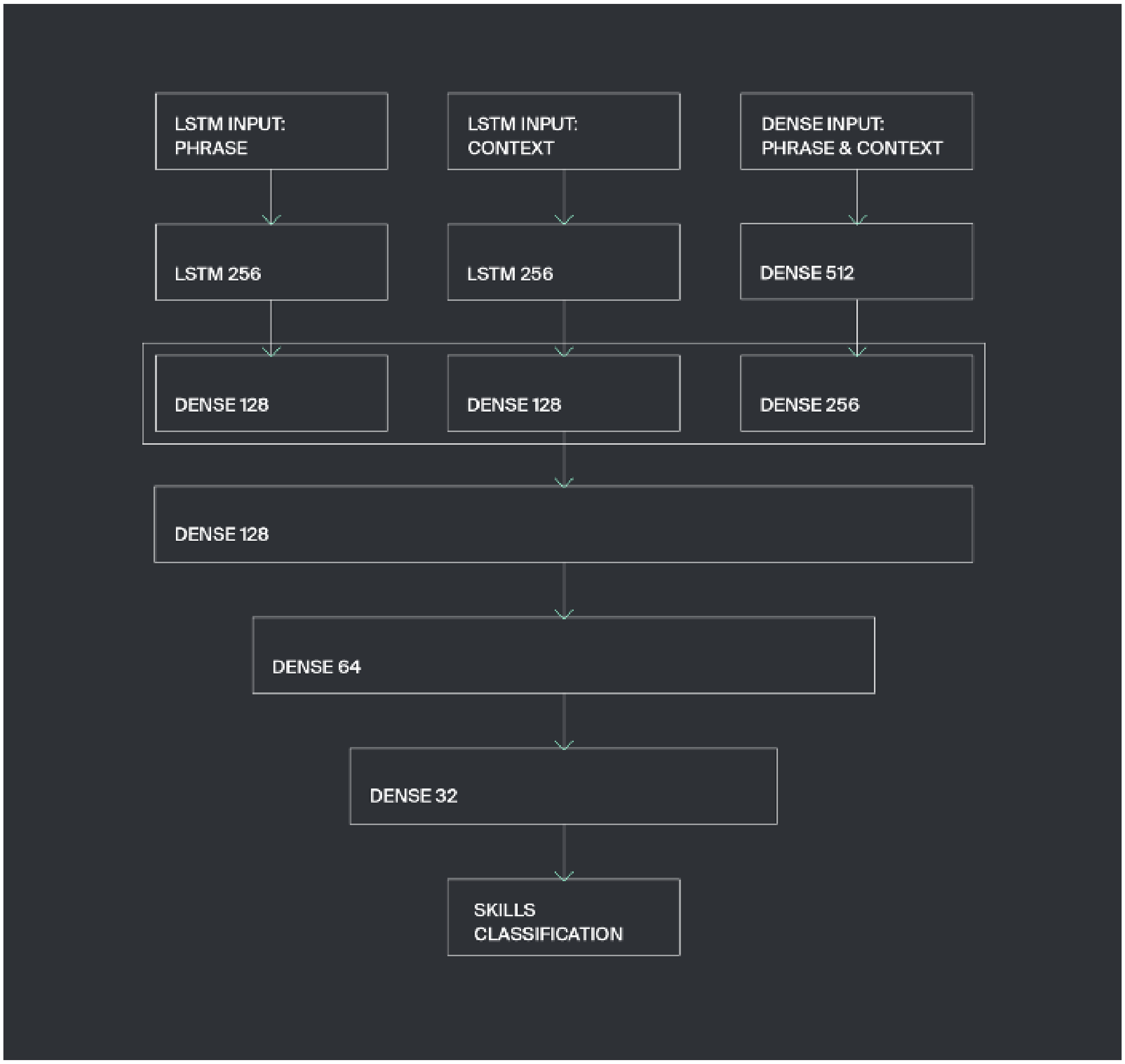


The second variable length vector brings the context structure information. For the given window size n we take n neighbouring words to the right and n words to the left of our candidate phrase, vector representations of these words are concatenated into the variable length vector and passed to the LSTM layer. We found that the optimal n=3.

4.4.3 The Model

The third input layer has fixed length and processes the vector with the general information about the candidate phrase and its context — coordinatewise maximum and minimum values of word vectors in the phrase and its context which, among the other information, represent the presence or absence of many binary features in the whole phrase.

We called this architecture SkillsExtractor, here it is.



The output is this model is a list of skills of an input paragraph. In our case the input paragraph is the speech to text version of the Candidate answer during the online interview.

4.4.4 Scoring

A score is calculated based on the detected set of skills mentioned in the interview and the requirements that are needed for the job based on the job description.

4.5 Platform Statistical Review Analysis

We will be merging all techniques discussed in this documentation to perform an analysis on the interviewee and providing useful feedback for the interviewer and the company which wants to hire the interviewee as a report on our platform.

Chapter 5

Smart Interview Platform

In this chapter we will be discussing the software developments phases and the building blocks of the platform as a website and how the user interacts with it. How the company and the interviewer can use this platform to maximize the chances of filtering an interviewee while also lowering cost and time consumed.

The structure of this chapter is as follows: **Section 1** discusses the model architecture. **Section 2** discussing the requirements electing phase. **Section 3** overviewing the product backlog. **Section 4** explains the release plan. **Section 5** shows the Entity Relationship Diagram (ERD). **Section 6** an Overview of the website. **Section 7** First User type (the interviewee) and user interaction. **Section 8** Second User type (interviewer and Employer) and interaction. **Section 9** The report Pipeline. **Section 10** explains the Flaskapi.

5.1 MVC Architecture

Model view controller (MVC) is a software design architecture which is designed commonly to reflect user interfaces in three interconnected elements. This is done to separate internal representations of information from the ways information is presented to and accepted from the user.

This Architecture is divided into three interconnected elements.

The central component of the architecture. **The Model** is the application's dynamic data structure, independent of the user interface. It directly manages the data, logic and rules of the application

Second element is **The View** which is a representation of information such as a chart, diagram or table. Multiple views of the same information are possible.

Third element is **The Controller** which accepts input and converts it to commands for the model or view.

In addition to dividing the application into these components, the model manages the data received from users imputed to the controller. The view presented the model in a particular format. The controller responds to the user inputs and performs interactions on the data model objects. Validating the input then passes the input to the model.

5.2 Requirements Electing Phase

We will be covering our requirements eliciting phase, and how we expressed and documented our requirements in a readable and understandable format for both normal users and the development teams, so as we are following agile principles and methodology we took into our consideration the Agile Principles, which are most related to requirements while building and documenting our requirements which are:

- Welcoming changing requirements.
- Encouraging face-to-face interaction.
- Focusing on simplicity.
- Delivering working software frequently.

As well as we took into our consideration that requirements will inevitably change, as stated in Scrum which is an agile methodology

Based on that we used "user stories" which is the most common form of expressing agile requirements

As an employer.

- I want to register on the platform so that I can have an account and all my data to be abstracted in one place.

- I want to login to the platform using my account to access all my data.

- I want my data to be secured so that no one can access my account and know my future plans.

- I want to add a vacancy so that I can fully fill my need for employees in a specific position.

- I want to describe the vacancy so that I can approach the details of the job requirements and the responsibilities to the user or give him some advice on how to make the interview more effective.

- I want to set a number of questions in the interview, so that I can get more flexibility in interview creation.

- I want to set skills to each question which the question targets so that I can determine the scoop of the question.

- I want to edit the vacancies details so that I can edit the mistakes and update the vacancies data based on the current situation.

- I want to send Invitations to registered or unregistered users by mail so that I can interview the recommendation network or the person who requests an interview from other platforms.

- I want to see the vacancy updates so that I can make a decision on beginning the filtration phase or sending more invitations.

- I want to list all vacancies so that I can get an overview of all vacancies progress.

- I want to test the interview so that I can determine the time of all questions passed or not.

- I want to show the report of every interviewed candidate so that I can filter the qualified candidates.

- I want to see the whole interview so that I can ensure the decision which I have taken.

- I want to search in the lists like question lists and vacancies lists so that I can retrieve the specific question easier.

- I want to select the number of questions to display in the list so that I can avoid confusion.

- I want to write notes in the vacancy so that I can approach specific ideas to the candidates.

- I want to see the state of each invitation so that I can observe the candidates progress.

- I want to have the ability to add, or delete questions in the vacancies so that I can update the vacancies based on the current situation.

As an interviewee.

- I want to register on the platform so that I can have an account and all my data to be abstracted in one place.
- I want to login to the platform using my account to access all my data.
- I want my data to be secured so that no one can access my account and know my future plans.
- I want to get a direct link to the interview when I get an invitation so that I can reach the interview as fast and easy as possible.
- I want to see the description of the vacancy before I start the interview so that I can determine if the position fits me or not.
- I want to know the allowed duration for each question so that I can manage my time.
- I want to know the question number that I am currently answering so that I can know how many questions remain.
- I want to see instructions about the interview process so that I can avoid the mistakes.
- I want to see the description of each question so that I can answer the specific answer which the interviewer wants.

- I want to get a success message after each question so that I can know if something wrong happened.

5.3 Prioritized Product Backlog

A product backlog is a prioritized list of work for the development team that is derived from the roadmap and its requirements. We will demonstrate the usage of the product backlog as it is a popular technique, which is critical to the agile methodology of Scrum.

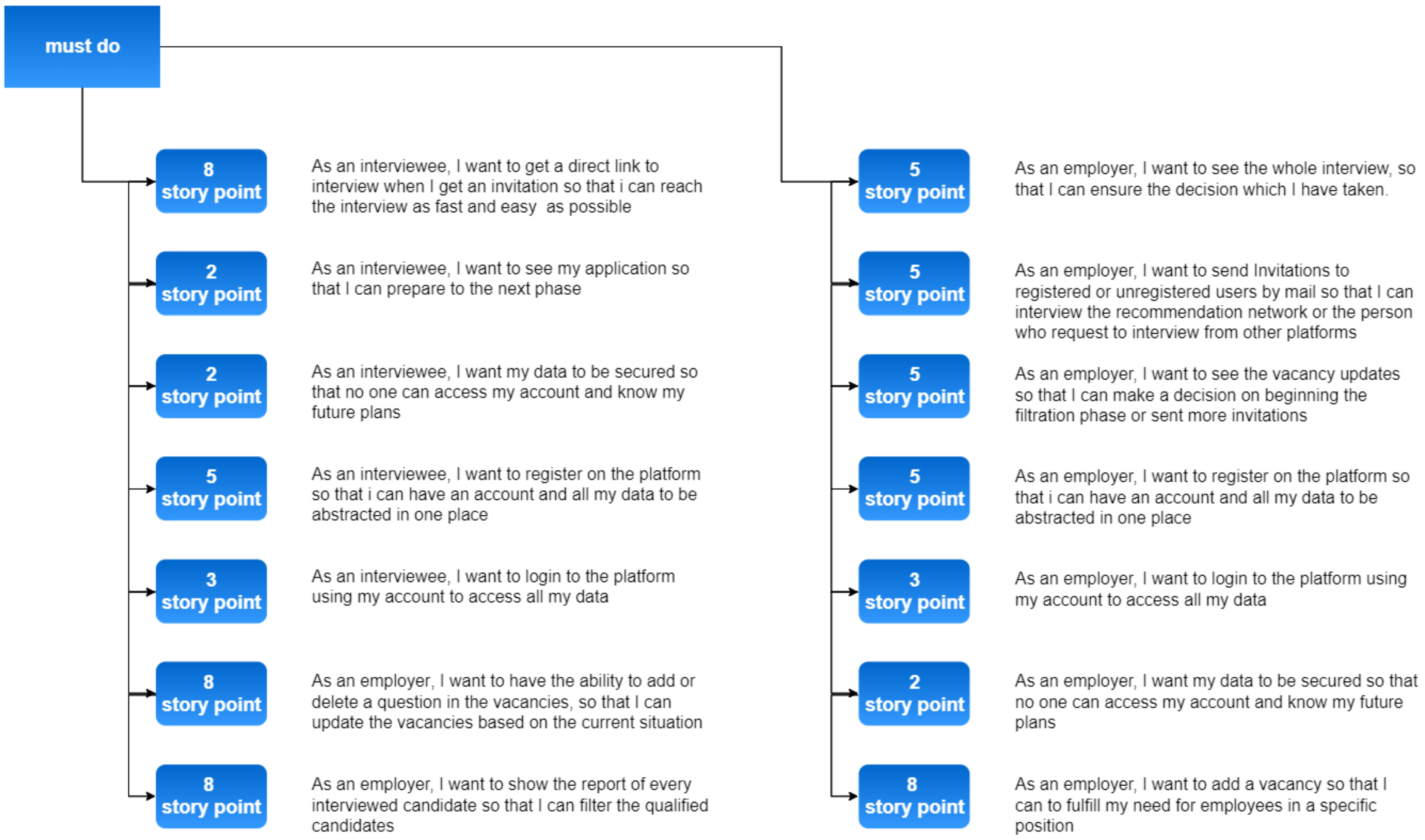
Basically, a product backlog is a list of software features which you and your team intend to develop. Actually, the product backlog begins as a big unsorted list of items to do then gets more refined over time.

Our product backlog consists mostly of user stories which we defined. The product backlog is a scrum technique, which also means it's an agile technique. That being the case, the product backlog should be dynamic, and focused on client interaction.

The initial user stories we wrote and gathered together and placed into the backlog.

There is no particular order to it, in fact the whole point of the backlog at the very beginning is that it's an unordered lump of all the things which would ideally be built into the end product.

Once we have our list, the next thing for us to do is to give a priority to each story. We will have the top priority user stories assigned a must do, the medium priority ones, A should do, and the low priority ones, A could do.



must do

8 story point

As an interviewee, I want to get a direct link to interview when I get an invitation so that i can reach the interview as fast and easy as possible

2 story point

As an interviewee, I want to see my application so that I can prepare to the next phase

2 story point

As an interviewee, I want my data to be secured so that no one can access my account and know my future plans

5 story point

As an interviewee, I want to register on the platform so that i can have an account and all my data to be abstracted in one place

3 story point

As an interviewee, I want to login to the platform using my account to access all my data

8 story point

As an employer, I want to have the ability to add or delete a question in the vacancies, so that I can update the vacancies based on the current situation

8 story point

As an employer, I want to show the report of every interviewed candidate so that I can filter the qualified candidates

5 story point

As an employer, I want to see the whole interview, so that I can ensure the decision which I have taken.

5 story point

As an employer, I want to send Invitations to registered or unregistered users by mail so that I can interview the recommendation network or the person who request to interview from other platforms

5 story point

As an employer, I want to see the vacancy updates so that I can make a decision on beginning the filtration phase or sent more invitations

5 story point

As an employer, I want to register on the platform so that i can have an account and all my data to be abstracted in one place

3 story point

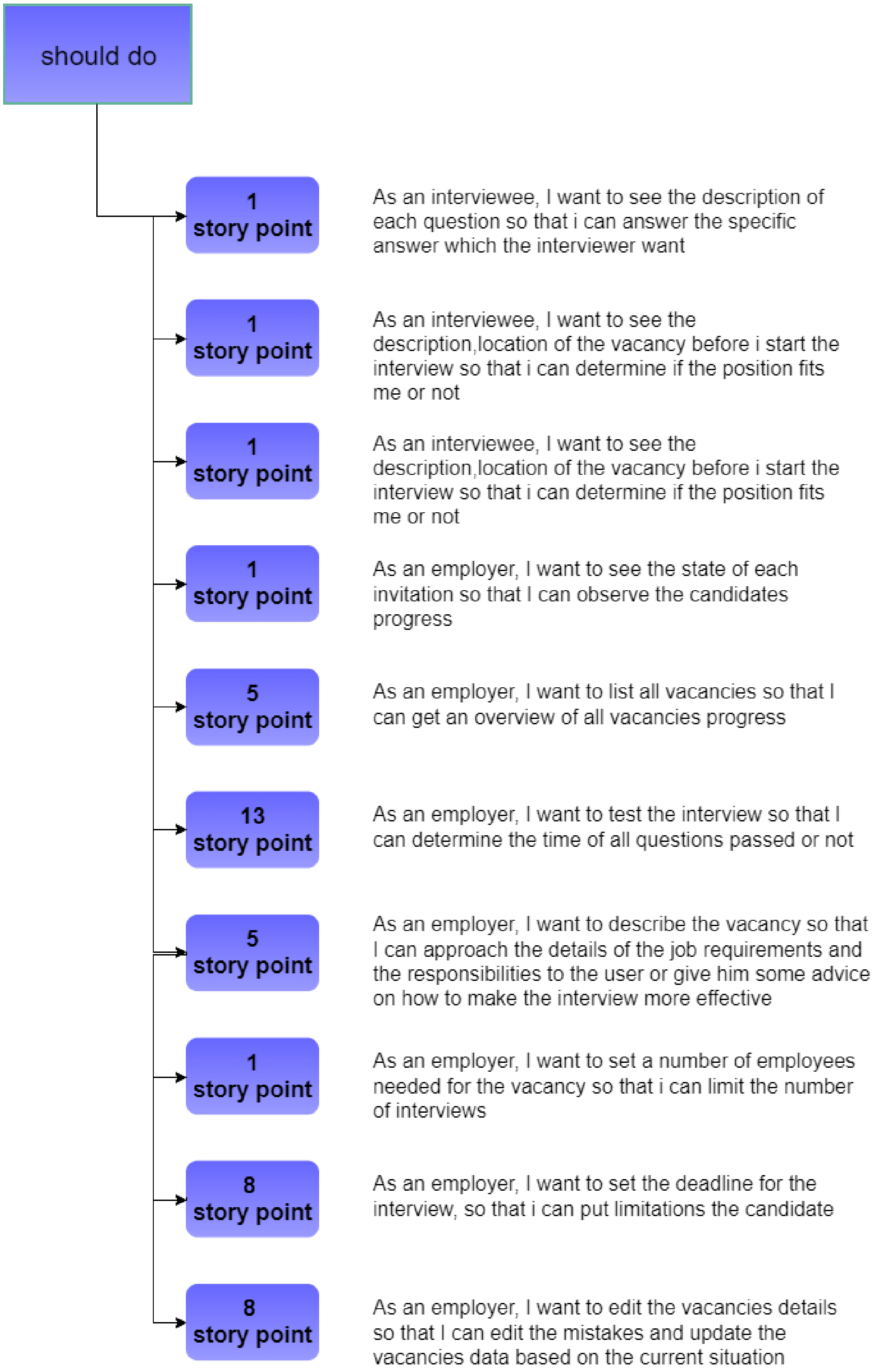
As an employer, I want to login to the platform using my account to access all my data

2 story point

As an employer, I want my data to be secured so that no one can access my account and know my future plans

8 story point

As an employer, I want to add a vacancy so that I can to fulfill my need for employees in a specific position



should do

1 story point

As an interviewee, I want to see the description of each question so that i can answer the specific answer which the interviewer want

1 story point

As an interviewee, I want to see the description,location of the vacancy before i start the interview so that i can determine if the position fits me or not

1 story point

As an interviewee, I want to see the description,location of the vacancy before i start the interview so that i can determine if the position fits me or not

1 story point

As an employer, I want to see the state of each invitation so that I can observe the candidates progress

5 story point

As an employer, I want to list all vacancies so that I can get an overview of all vacancies progress

13 story point

As an employer, I want to test the interview so that I can determine the time of all questions passed or not

5 story point

As an employer, I want to describe the vacancy so that I can approach the details of the job requirements and the responsibilities to the user or give him some advice on how to make the interview more effective

1 story point

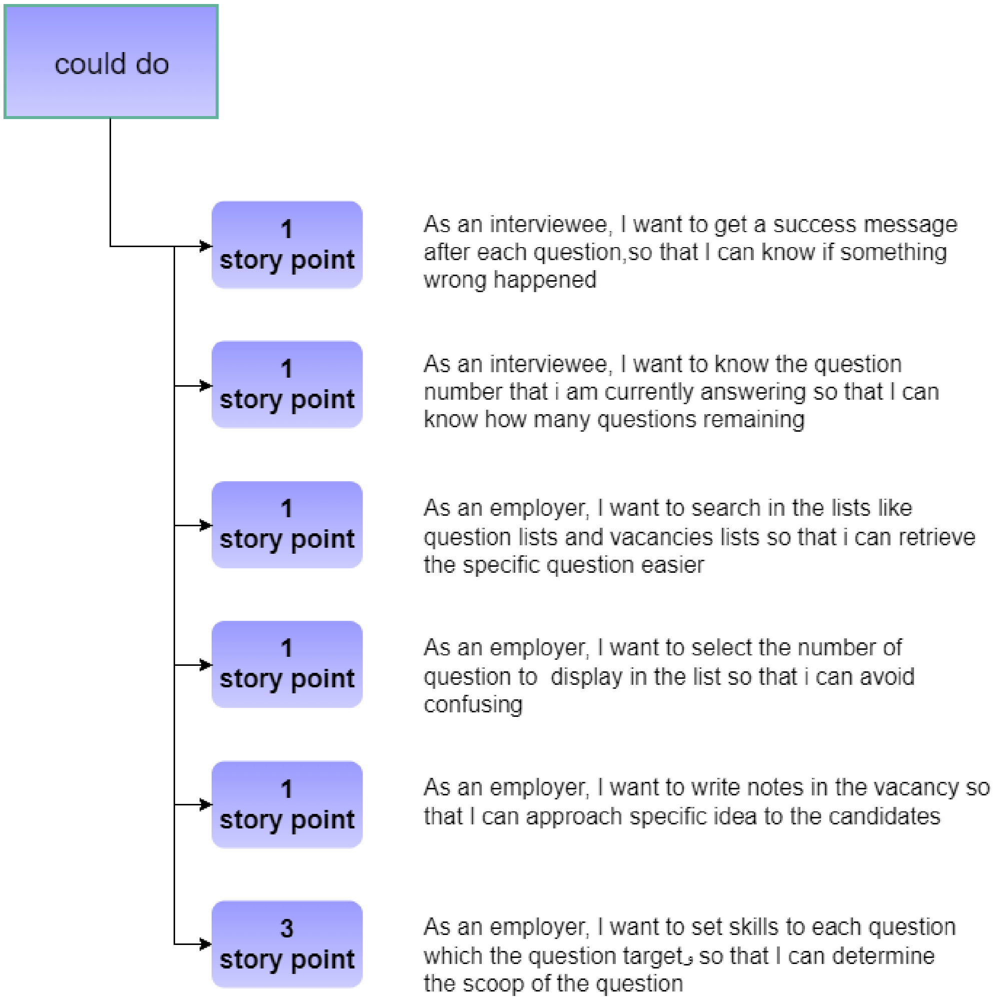
As an employer, I want to set a number of employees needed for the vacancy so that i can limit the number of interviews

8 story point

As an employer, I want to set the deadline for the interview, so that i can put limitations the candidate

8 story point

As an employer, I want to edit the vacancies details so that I can edit the mistakes and update the vacancies data based on the current situation



As an interviewee, I want to get a success message after each question,so that I can know if something wrong happened

As an interviewee, I want to know the question number that i am currently answering so that I can know how many questions remaining

As an employer, I want to search in the lists like question lists and vacancies lists so that i can retrieve the specific question easier

As an employer, I want to select the number of question to display in the list so that i can avoid confusing

As an employer, I want to write notes in the vacancy so that I can approach specific idea to the candidates

As an employer, I want to set skills to each question which the question target، so that I can determine the scoop of the question

Then the next step is to start planning for our project using these priorities as our reference point. Where user stories are taken from the backlog and placed into chunks of work, which should be done in a certain time interval called sprints.

But first we need to determine our "story points" and "velocity estimate”, to use these estimates into a tangible plan

- **Story points:**

It's hard to make reliably accurate estimates especially when we make those estimates far into the future. The way Story Points try to solve this problem is to eliminate time as the unit of measurement for estimating work.

Instead, Story Points are used when estimating required work to be done. Story points are unit-less and relative. The point is, that they allow the developers to move away from trying to estimate the exact amount of time some work will take and towards how long it will take in relation to other pieces of work.

So story points are built in a relative manner where each user story is assigned a number relative to another user story.

- **Velocity estimate:**

Velocity is the amount of work done within a period of time you've spent doing that work. Conventionally, in the agile community, the team's velocity is measured using the number of story points completed for the user stories done within the duration of a sprint.

So our software project velocity is determined as 15 story points after experimenting this in the first few sprints.

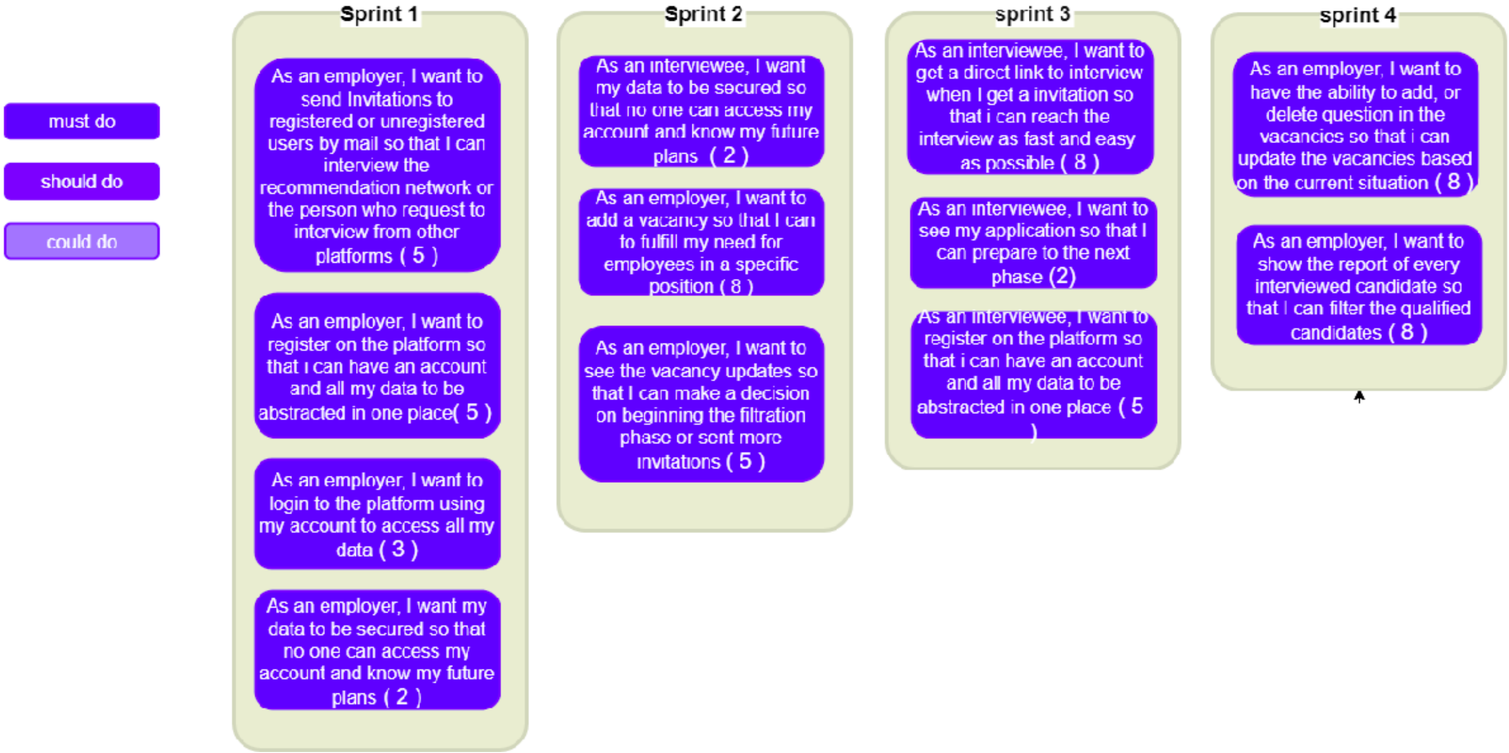
5.4 Release Plan

In Scrum, we have time boxed iterations called sprints. Release planning is used to determine which user stories should be completed and released by the end of each sprint.

For the next sprints, the user stories come from the product backlog, and are scheduled into the upcoming sprints based on their priority.

We started with release planning to place user stories amongst sprints, before starting the first sprint.

After prioritizing our user stories what you end up with is a good idea of what we view as a priority, before moving into development. Our prioritized user stories are then taken from the top of the priority list and distributed throughout your release plan into sprints. Until each sprint is filled with an appropriate amount of work.



must do

should do

could do

Sprint 1

As an employer, I want to send Invitations to registered or unregistered users by mail so that I can interview the recommendation network or the person who request to interview from other platforms (5)

As an employer, I want to register on the platform so that ı can have an account and all my data to be abstracted in one place(5)

As an employer, I want to login to the platform using my account to access all my data (3)

As an employer, I want my data to be secured so that no one can access my account and know my future plans (2)

Sprint 2

As an interviewee, I want my data to be secured so that no one can access my account and know my future plans (2)

As an employer, I want to add a vacancy so that I can to fulfill my need for employees in a specific position (8)

As an employer, I want to see the vacancy updates so that I can make a decision on beginning the filtration phase or sent more invitations (5)

sprint 3

As an interviewee, I want to get a direct link to interview when I get a invitation so that i can reach the interview as fast and easy as possible (8)

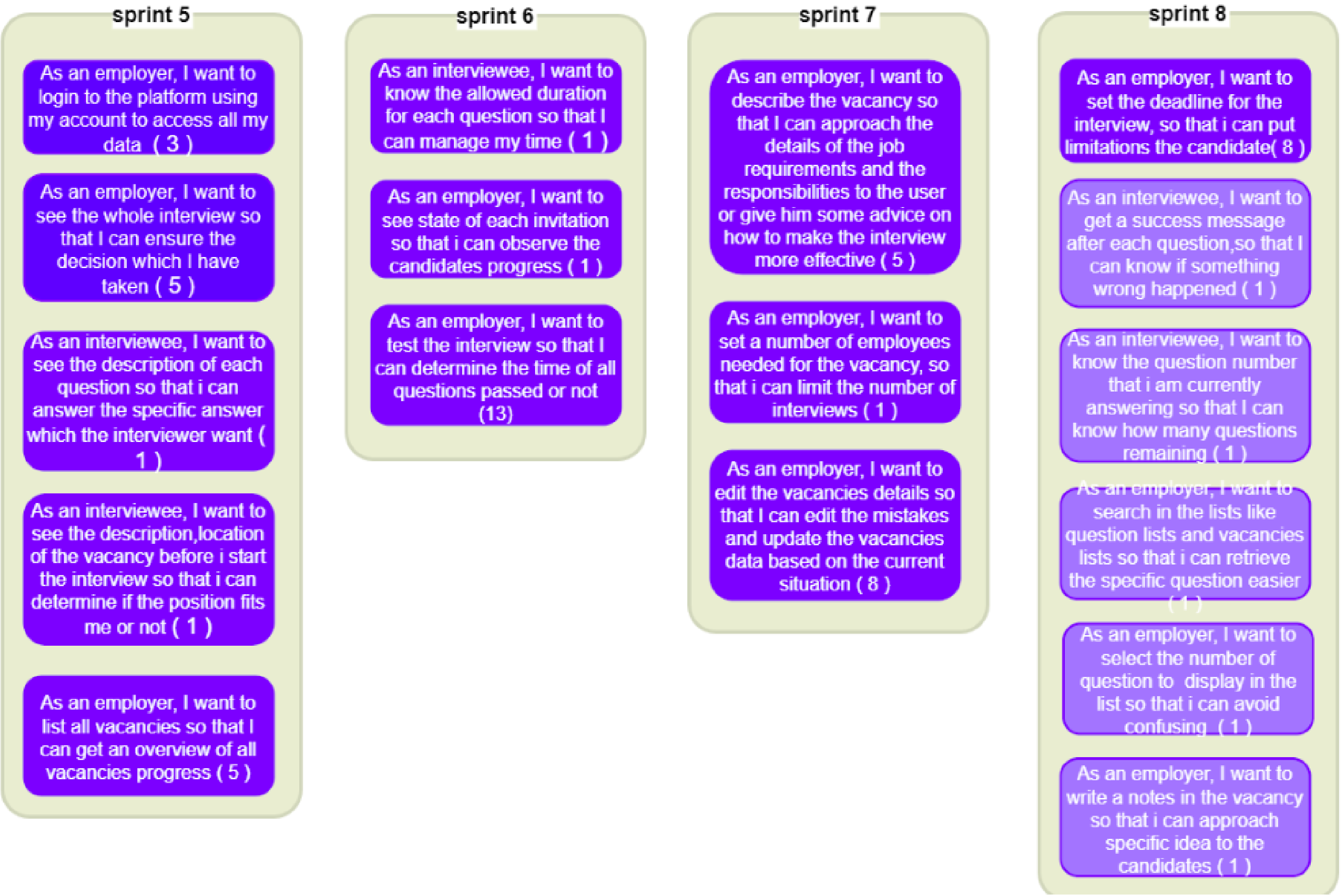
As an interviewee, I want to see my application so that I can prepare to the next phase (2)

As an interviewee, I want to register on the platform so that i can have an account and all my data to be abstracted in one place (5)

sprint 4

As an employer, I want to have the ability to add, or delete question in the vacancies so that i can update the vacancies based on the current situation (8)

As an employer, I want to show the report of every interviewed candidate so that I can filter the qualified candidates (8)



sprint 5

As an employer, I want to login to the platform using my account to access all my data (3)

As an employer, I want to see the whole interview so that I can ensure the decision which I have taken (5)

As an interviewee, I want to see the description of each question so that i can answer the specific answer which the interviewer want (1)

As an interviewee, I want to see the description,location of the vacancy before i start the interview so that i can determine if the position fits me or not (1)

As an employer, I want to list all vacancies so that I can get an overview of all vacancies progress (5)

sprint 6

As an interviewee, I want to know the allowed duration for each question so that I can manage my time (1)

As an employer, I want to see state of each invitation so that i can observe the candidates progress (1)

As an employer, I want to test the interview so that I can determine the time of all questions passed or not (13)

sprint 7

As an employer, I want to describe the vacancy so that I can approach the details of the job requirements and the responsibilities to the user or give him some advice on how to make the interview more effective (5)

As an employer, I want to set a number of employees needed for the vacancy, so that i can limit the number of interviews (1)

As an employer, I want to edit the vacancies details so that I can edit the mistakes and update the vacancies data based on the current situation (8)

sprint 8

As an employer, I want to set the deadline for the interview, so that i can put limitations the candidate(8)

As an interviewee, I want to get a success message after each question,so that I can know if something wrong happened (1)

As an interviewee, I want to know the question number that i am currently answering so that I can know how many questions remaining (1)

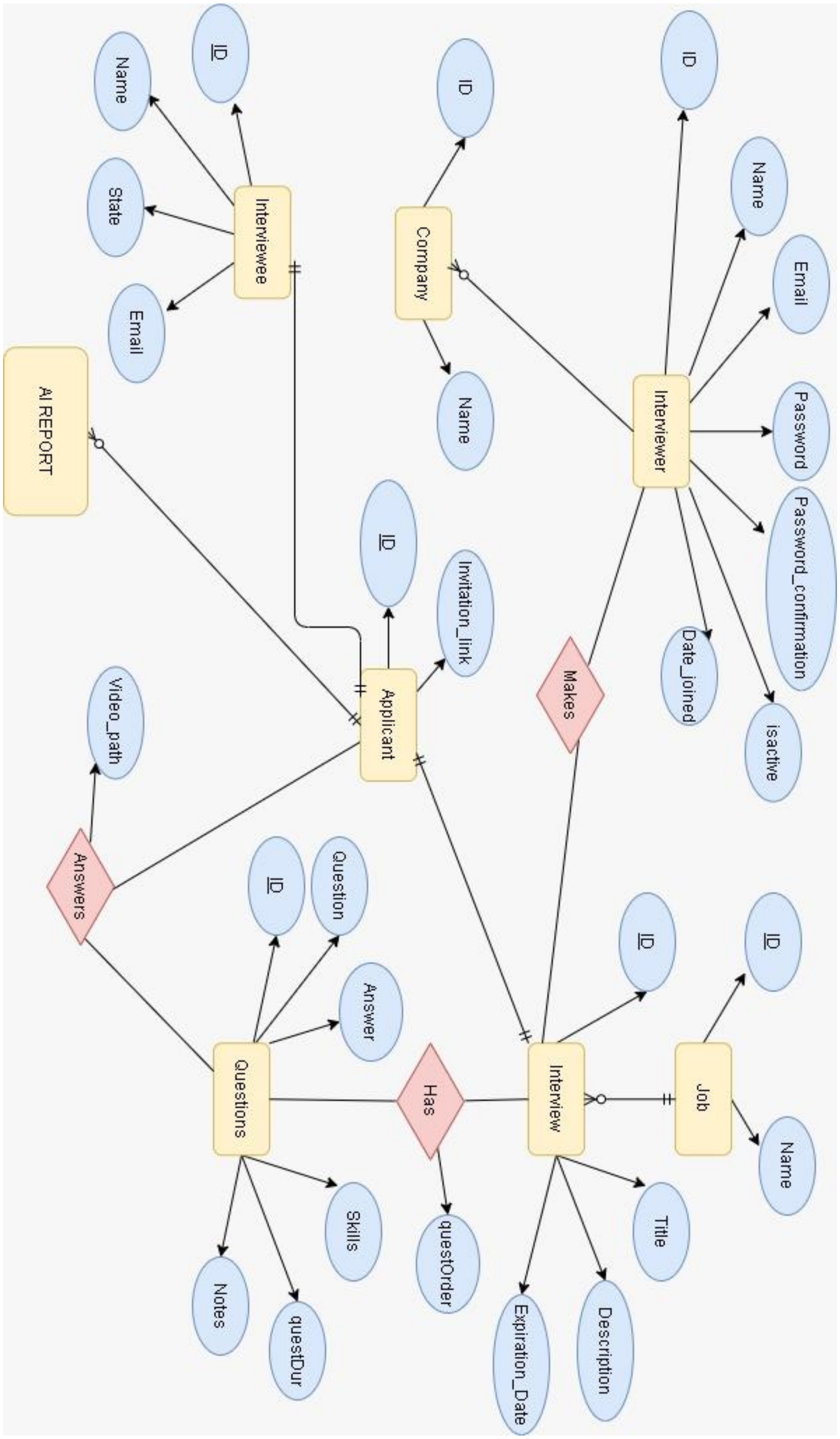
As an employer, I want to search in the lists like question lists and vacancies lists so that i can retrieve the specific question easier (1)

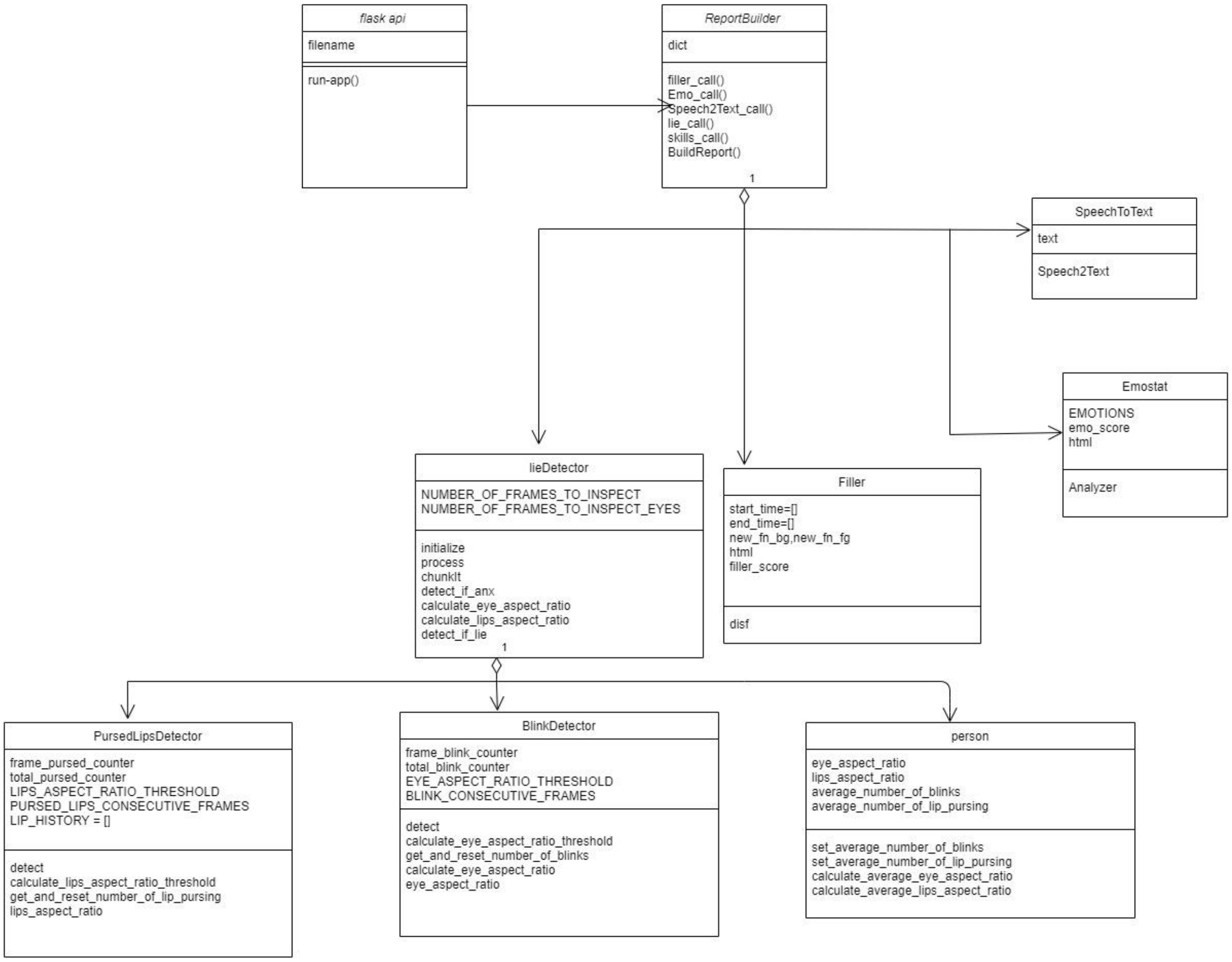
As an employer, I want to select the number of question to display in the list so that i can avoid confusing (1)

As an employer, I want to write a notes in the vacancy so that i can approach specific idea to the candidates (1)

5.4 Enhanced ERD and Class Diagram

Enhanced entity relationship diagram or Extended entity relationship diagram, this was developed to reflect precisely the properties and constraints that are found in our platform databases.

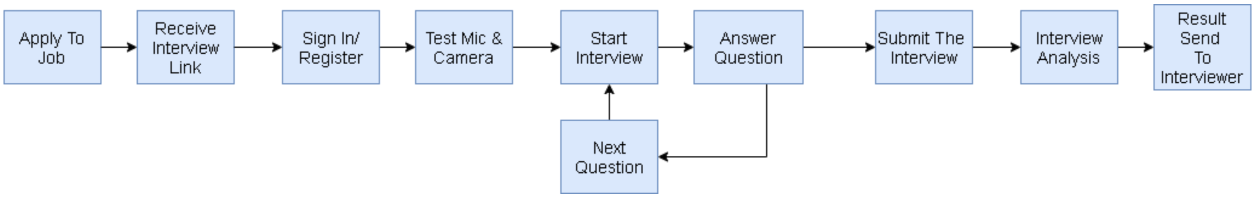

Interviewee
ID
Name
State
Email
AI REPORT
Company
ID
Name
Interviewer
ID
Name
Email
Password
Password_confirmation
Date_joined
isactive
Makes
Applicant
ID
Invitation_link
Answers
Video_path
Questions
ID
Question
Answer
Skills
Notes
questDur
Has
questOrder
Interview
ID
Title
Description
Expiration_Date
Job
ID
Name



Intelligent video interview mainly targets two types of users, the Employees who are trying the best candidate for the job, and the candidates who are trying to prove that they are the most fitting for the job. And we have made sure that our website provides those two objectives without any complications for any of them.

We'll discuss the scenarios for both types of users and later in the documentation we will discuss all the processes that happen in the background.

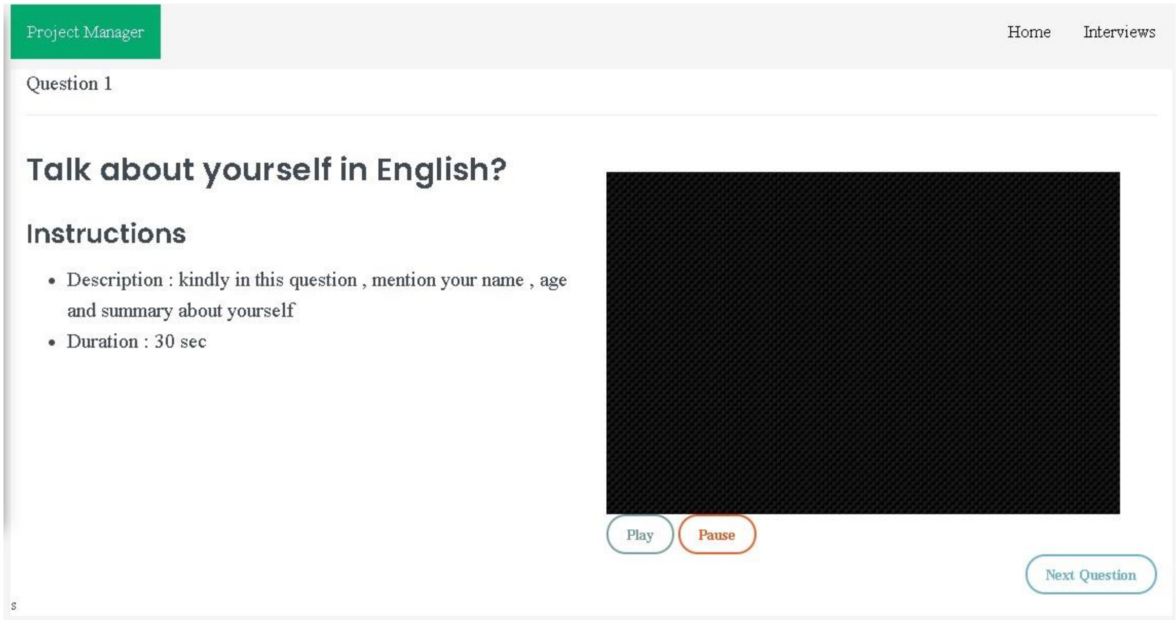
5.7 First User Type (Interviewee)



The Candidate first applies to the job, and if the Employer wishes for him to continue to the Online Interview, The candidate will be sent a link where he can proceed to the interview page

After signing in or registering to our website he is then greeted to the website and is provided by instructions/hints that they should follow for optimal results during the interview, for example to sit in an environment that has good lighting, to have a good microphone, to have a working webcam...etc. and the candidate can then do a test question to see if everything is working correctly before proceeding to the actual interview.

Once the candidate proceeds to the actual interview page he will start going through a set of questions set by the Employer and answering them via his webcam and microphone one by one.



Question 1

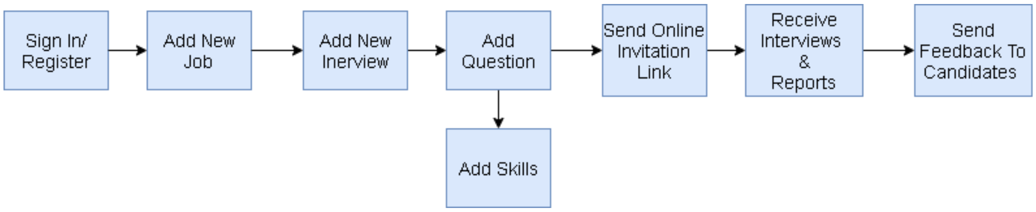
Talk about yourself in English?

Instructions

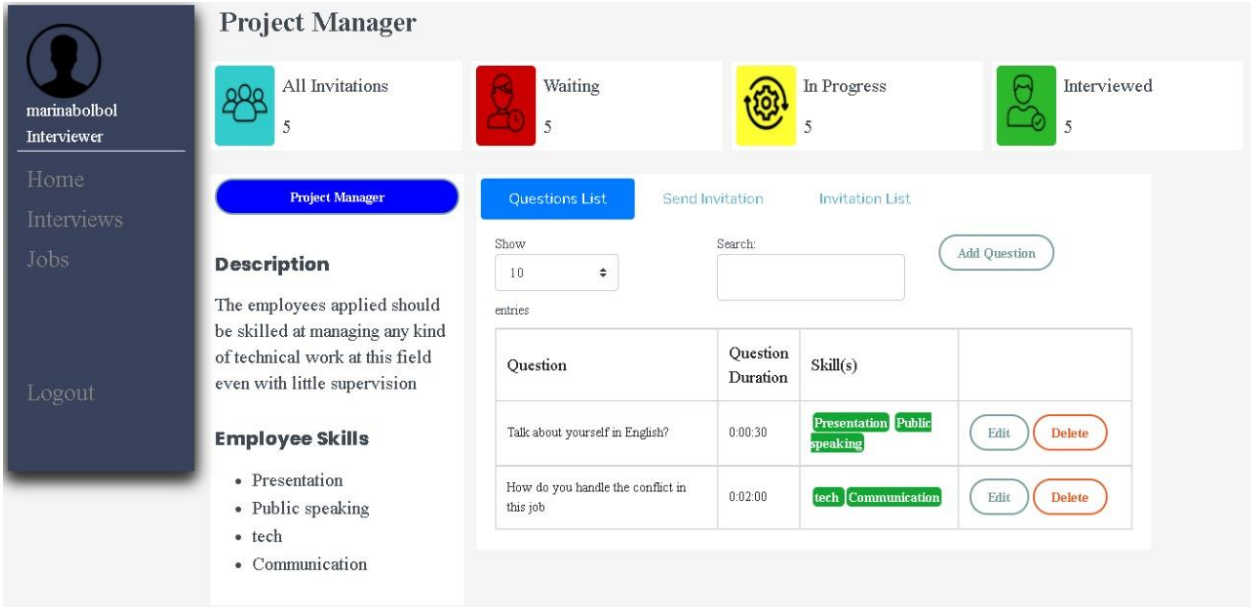
- Description : kindly in this question , mention your name , age and summary about yourself
- Duration : 30 sec

Once done he will be forwarded to a page telling him that his interview is done and he should keep checking his email for the interview feedback by the employer.

5.8 Second User Type (Interviewer, Employer)


Sign In/ Register
Add New Job
Add New Inerview
Add Question
Add Skills
Send Online Invitation Link
Receive Interviews & Reports
Send Feedback To Candidates

After the Employer signs in/register to the website. He add new vacancy (new job) then add new vacancy's interview including the questions that the candidate needs to answer during the interview, The Employer can also add skills to each questions for easier and more accurate report results.



Project Manager

marinabolbol
Interviewer

All Invitations 5

Waiting 5

In Progress 5

Interviewed 5

Description

The employees applied should be skilled at managing any kind of technical work at this field even with little supervision

Employee Skills

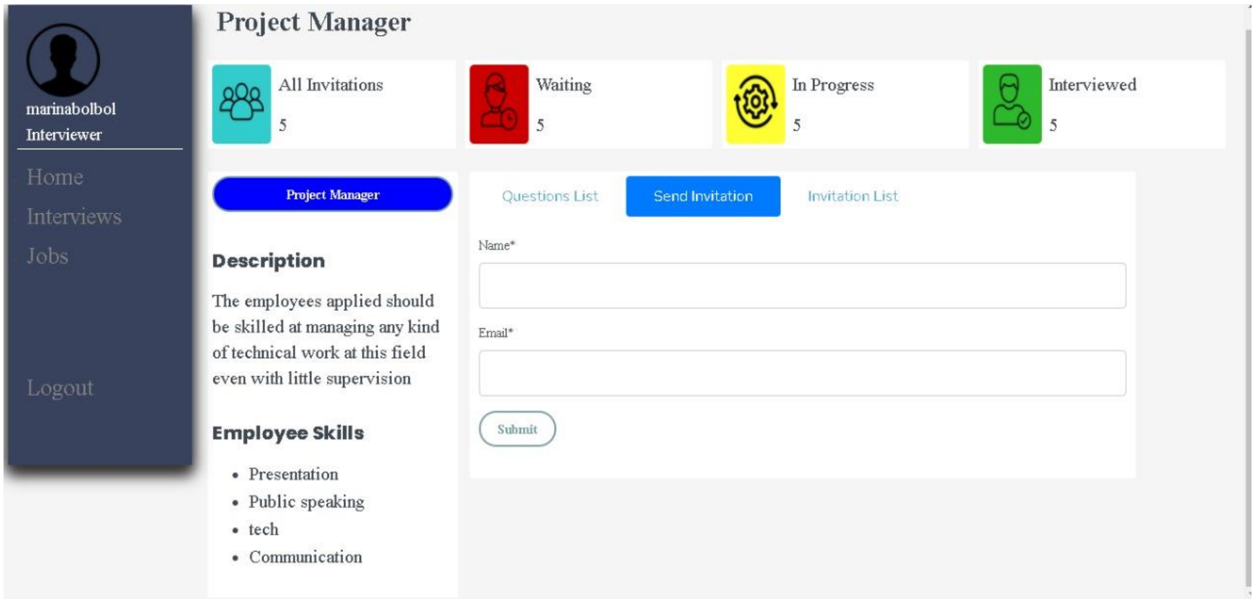
- Presentation
- Public speaking
- tech
- Communication

Show 10 entries

Search:

Question	Question Duration	Skill(s)	
Talk about yourself in English?	0:00:30	Presentation Public speaking	Edit Delete
How do you handle the conflict in this job	0:02:00	tech Communication	Edit Delete

The employer can send a link that has a unique ID linked to a candidate's interview page where the candidate can sign in/ register and start taking the steps previously mentioned to submit his interview.



Project Manager

marinabolbol
Interviewer

All Invitations 5

Waiting 5

In Progress 5

Interviewed 5

Description

The employees applied should be skilled at managing any kind of technical work at this field even with little supervision

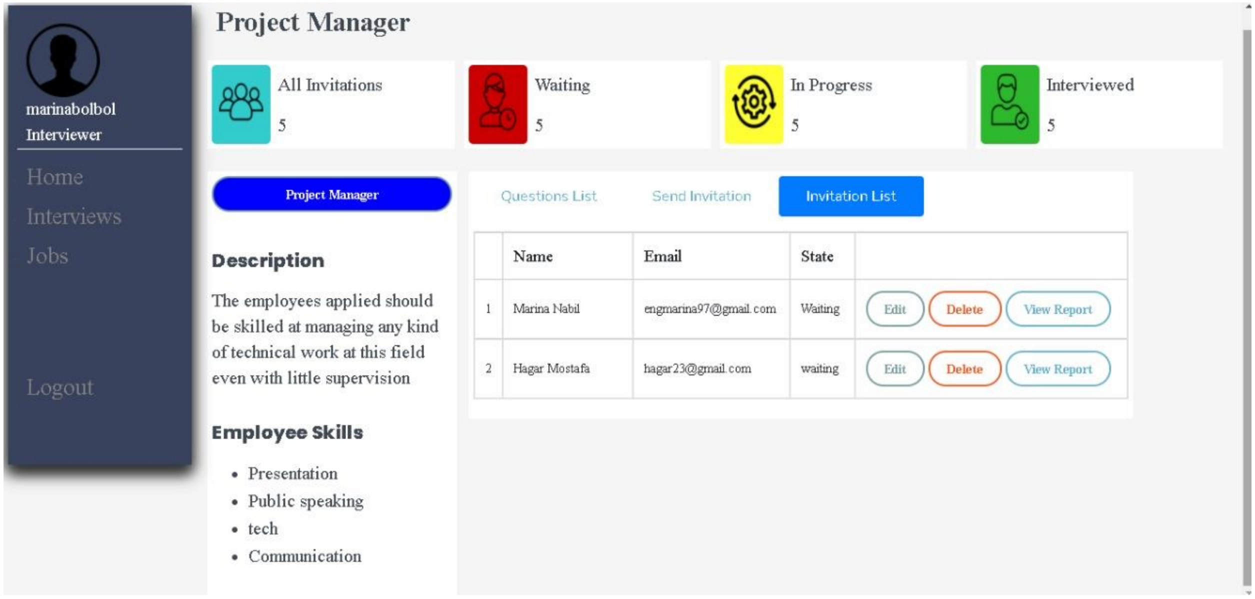
Employee Skills

- Presentation
- Public speaking
- tech
- Communication

Name*

Email*

Once the candidate submits his interview, it is then analyzed and a report is built using our AI, and then a list of reports appear in the Employee's dashboard, arranged from best candidate to worst candidate based on a scoring system that we will talk about in details later



Project Manager

All Invitations	Waiting	In Progress	Interviewed
5	5	5	5

Description

The employees applied should be skilled at managing any kind of technical work at this field even with little supervision

Employee Skills

- Presentation
- Public speaking
- tech
- Communication

	Name	Email	State	
1	Marina Nabil	engmarina97@gmail.com	Waiting	Edit Delete View Report
2	Hagar Mostafa	hagar23@gmail.com	waiting	Edit Delete View Report

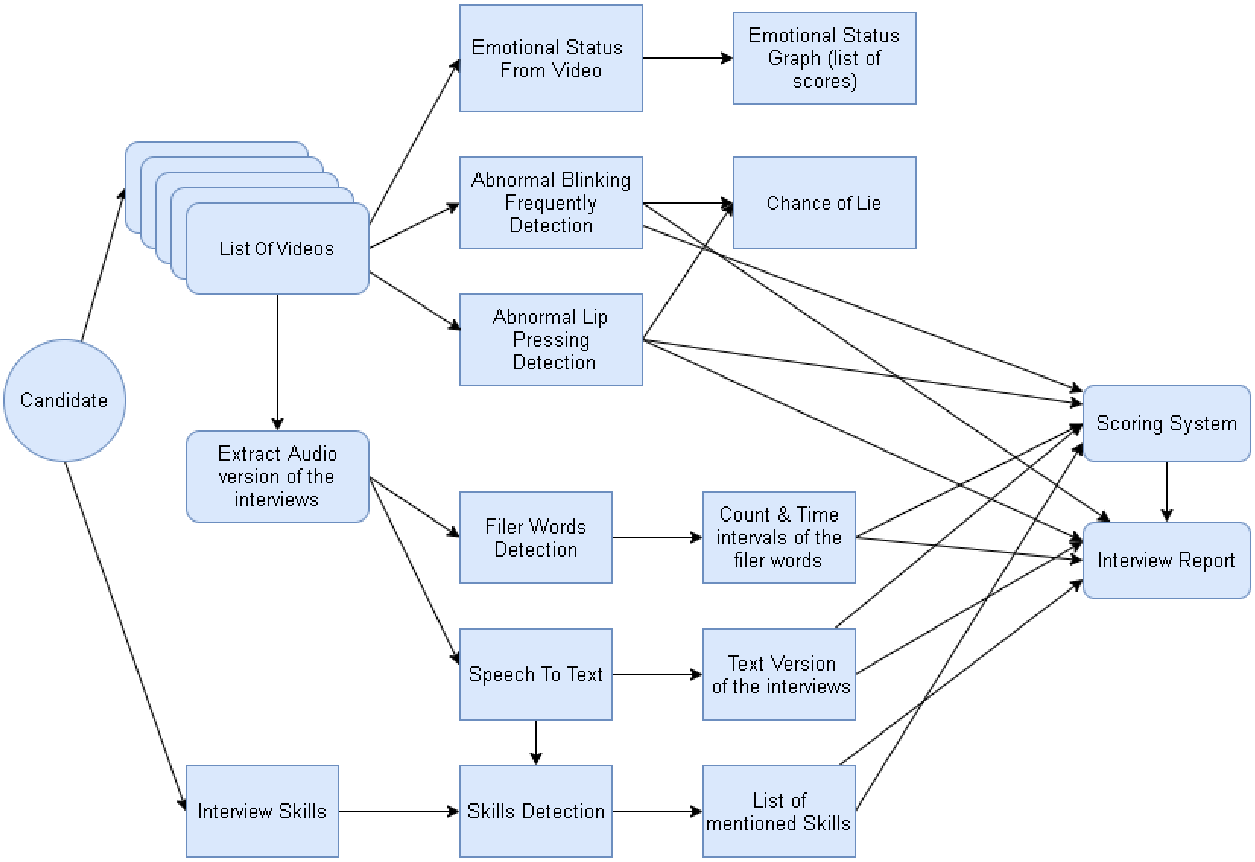
The employer can then click on each report and see a list of questions, the scores that the candidate has gotten on each question, and a summarized version of the mistakes that the candidate has done on each question.

The employer can then click on each question to see an in depth report of the question, he can see the video and a bunch of other things that we will talk about in detail later.

The employer then sends feedback emails to the candidates that he wishes they continue to the next recruiting phase.

5.9 The Report Building Pipeline

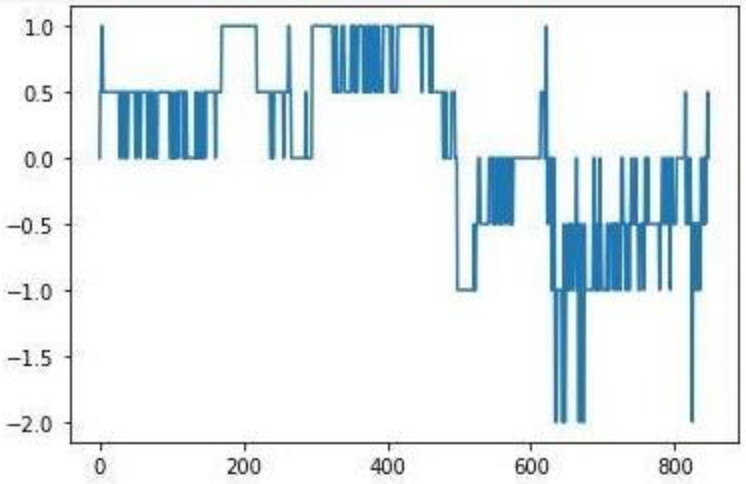
So far we have talked about building a report but we haven't talked about what that report will contain and how exactly it will be built.

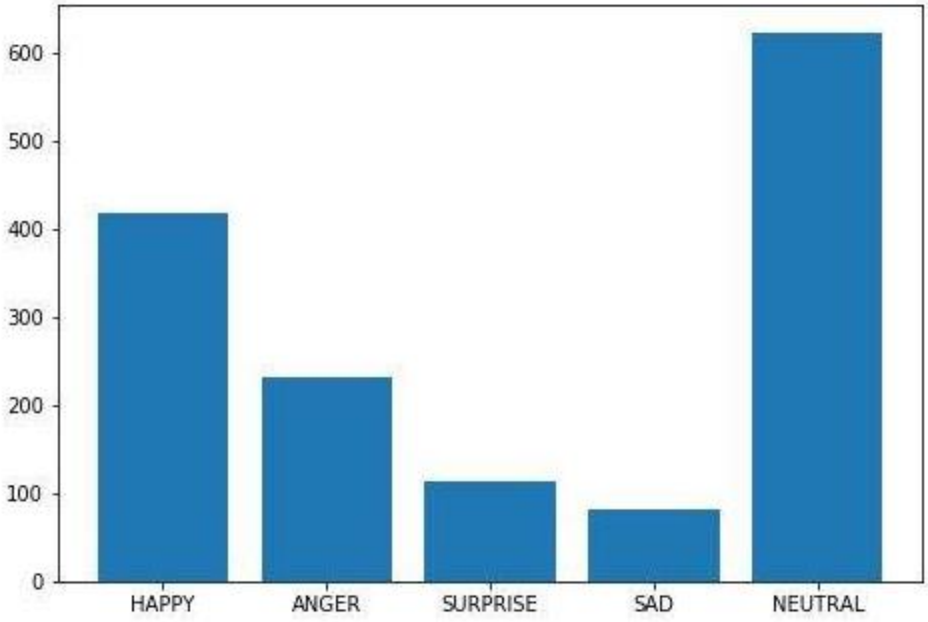

Emotional Status From Video
Emotional Status Graph (list of scores)
Abnormal Blinking Frequently Detection
Chance of Lie
List Of Videos
Abnormal Lip Pressing Detection
Candidate
Scoring System
Extract Audio version of the interviews
Filer Words Detection
Count & Time intervals of the filer words
Interview Report
Speech To Text
Text Version of the interviews
Interview Skills
Skills Detection
List of mentioned Skills

The above graph shows an overview of how the report is built. We will talk generally about how the report is built and then we will go even more in depth about how each model was built.

It starts by the Candidate uploading and submitting his CV and his Interview videos.

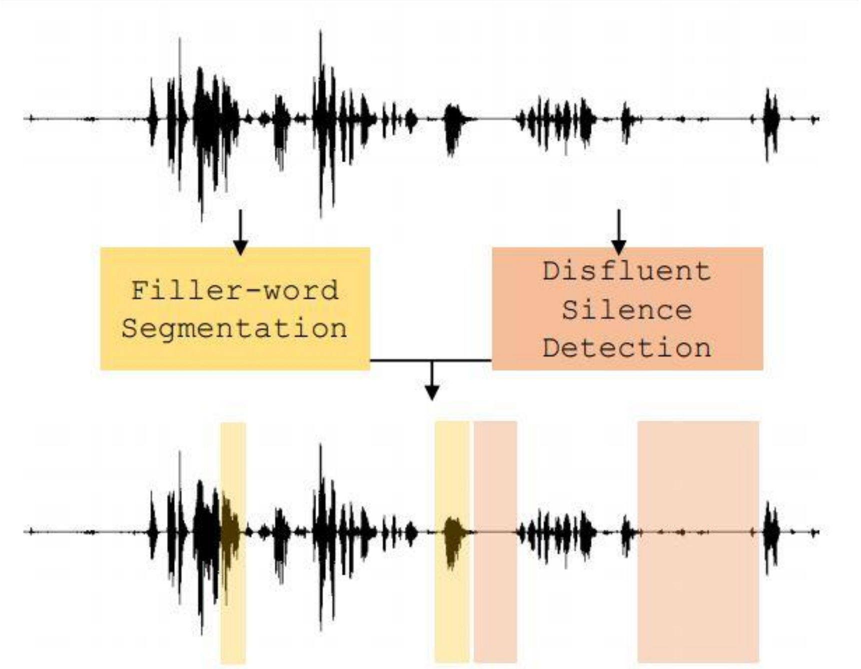
Each video goes through an emotional status model that returns a list of scores (a number for each frame) that is numbered between -1 and 1, -1 being very visibly mad/sad, and 1 being very visibly happy or excited. And this list of numbers is then graphed on a timeline to show the candidate’s emotional status throughout the video.





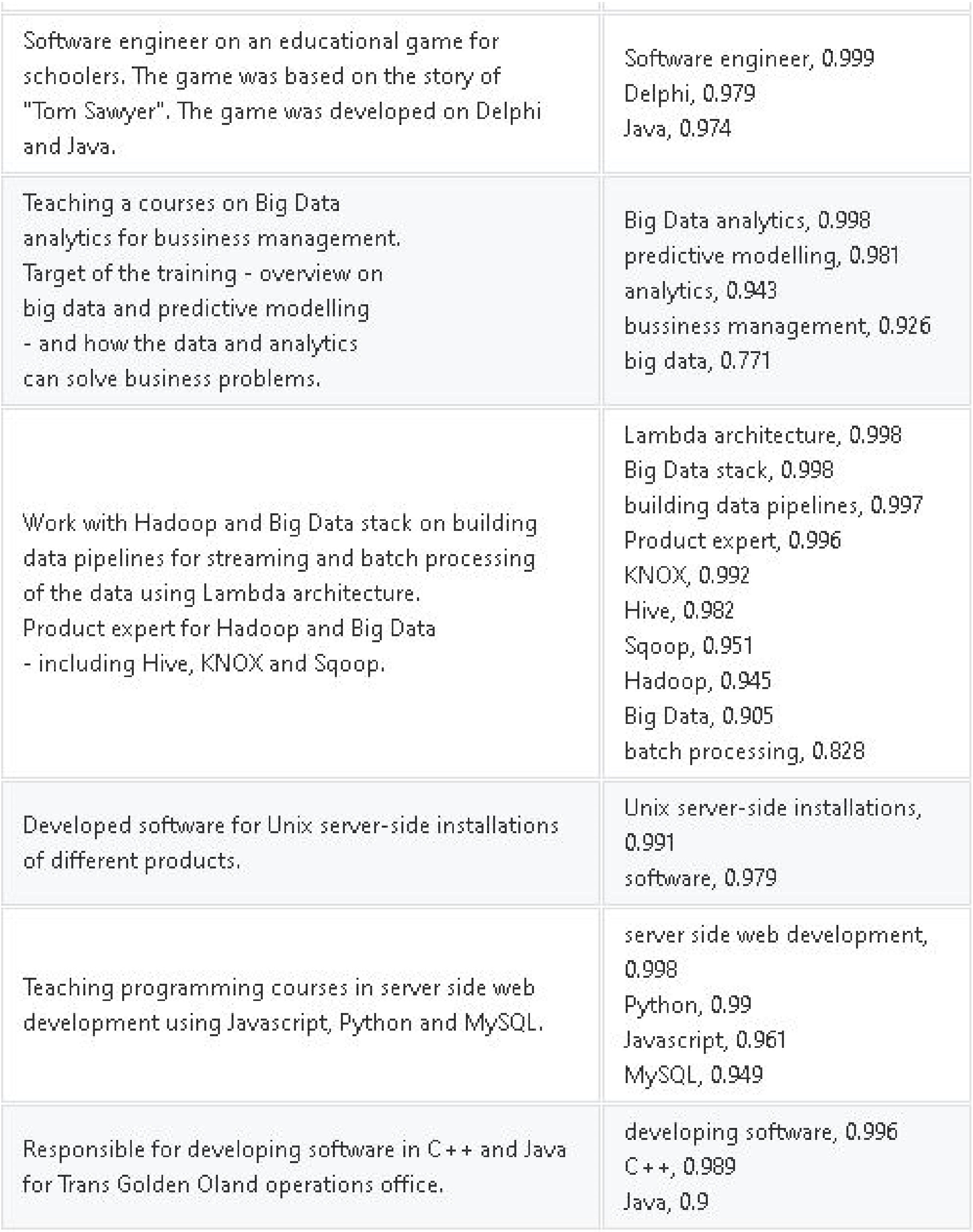
The video also goes through Abnormal Blinking Frequency Detection & Abnormal Lip Pressing detection models, that both detect whether the candidates blinks more than usual or presses too much on his lips during the video, and that is added to the report, if the candidate happens to do both during the video, this can indicate that there is a chance of the candidate lying while answering this question. And that is also added to the report to notify the employer.

A wav file is extracted from the video and goes through two models.
The first one is the Filler word detection model, this model detects all the Umms & Aahhs that the candidate says during the interview, and it returns its count, and the starting and ending timestamp of each umm, those timestamps are then visualized in our report to make it easier for the employer to find patterns, and detect the questions that the candidate has had trouble answering.



The second model that our Wav file passes through is a simple Speech to Text file, A text version of the candidate's answer is added to the report, that makes it easier for the employer to scrim through answers rather than having to watch the entire video to know what exactly the candidate's answer is.

The extracted text is also passed through a model that extracts skill from unstructured text which will be highlighted in the text version of the interview.



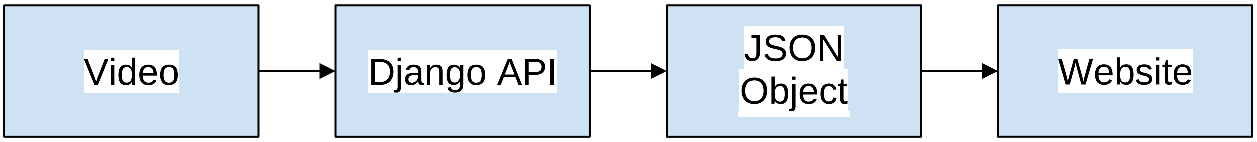
Software engineer on an educational game for schoolers. The game was based on the story of "Tom Sawyer". The game was developed on Delphi and Java.	Software engineer, 0.999 Delphi, 0.979 Java, 0.974
Teaching a courses on Big Data analytics for bussiness management. Target of the training - overview on big data and predictive modelling - and how the data and analytics can solve business problems.	Big Data analytics, 0.998 predictive modelling, 0.981 analytics, 0.943 bussiness management, 0.926 big data, 0.771
Work with Hadoop and Big Data stack on building data pipelines for streaming and batch processing of the data using Lambda architecture. Product expert for Hadoop and Big Data - including Hive, KNOX and Sqoop.	Lambda architecture, 0.998 Big Data stack, 0.998 building data pipelines, 0.997 Product expert, 0.996 KNOX, 0.992 Hive, 0.982 Sqoop, 0.951 Hadoop, 0.945 Big Data, 0.905 batch processing, 0.828
Developed software for Unix server-side installations of different products.	Unix server-side installations, 0.991 software, 0.979
Teaching programming courses in server side web development using Javascript, Python and MySQL.	server side web development, 0.998 Python, 0.99 Javascript, 0.961 MySQL, 0.949
Responsible for developing software in C++ and Java for Trans Golden Oland operations office.	developing software, 0.996 C++, 0.989 Java, 0.9

The CV is also passed through the model, and parsed so that the Employer can see a much summarized version of the CV on the list of interview reports in fig(x).

5.10 Django (Web framework)

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It’s free and open source. Django was designed to help developers take applications from concept to completion as quickly as possible. Django takes security seriously and helps developers avoid many common security mistakes. Some of the busiest sites on the Web leverage Django’s ability to quickly and flexibly scale.

We used Django to wrap our models in an API that the website can easily communicate with.



Django is a restful API where its end points are:

- Post generate report :
    - Description : this end point is used to generate the interviewee analysis and feedback
- Link : post("https://localhost/{vidoe_path}");
- Parameters :

Parameter	Type	Required	Description
video path	string	Yes	The path of the video

- Response model :
    - "fillerGraph": HTML representation of the emotion graph,
    - "fillerCount": Number of Disfluencies,
    - "fillerScore": Calculated score of disfluencies,
    - "text": Text Version of the interview,
    - "fillerGraph": HTML representation of the disfluencies graph,
    - "emotionScore": Calculated score of emotions,
    - "noOfBlinks": Number of blinks,
    - "noOfLipPursing": Number of lip pursing,
    - "blinksPerSecond": Ratio of Blinks Per Second,
    - "anxBlinks": Calculated score of anxiety based on blinks,
    - "anxLips": Calculated score of anxiety based on lips,
    - "anxScore": Calculated score of anxiety based on both blinks and lips,
    - "fillerGraph": Skills/Topics mentioned in the answer,

"totalScore": Total Calculated score

Experiments, Results and Discussion

This chapter verifies the dataset, experiments that took place and conclude the results.

The structure of this chapter is as follows: **Section 1** talks briefly about the dataset. **Section2** explains the experiments that took place. **Section 3** represents the result of our platform.

6.1 Dataset

We mainly used 3 deep learning models in our project, we'll talk a little bit about our data that we used to train those models in little details in this section:

- **Emotional status dataset**

As we have mentioned before, our emotional status uses mini-xception model which was trained on a FER dataset.

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

train.csv contains two columns, "emotion" and "pixels". The "emotion" column contains a numeric code ranging from 0 to 6, inclusive, for the emotion that is present in the image. The "pixels" column contains a string surrounded in quotes for each image. The contents of this string are space-separated pixel values in row major order. test.csv contains only the "pixels" column and your task is to predict the emotion column.

The training set consists of 28,709 examples. The public test set used for the leaderboard consists of 3,589 examples.

This dataset was prepared by Pierre-Luc Carrier and Aaron Courville, as part of an ongoing research project. They have graciously provided the workshop organizers with a preliminary version of their dataset to use for this contest.

- **Disfluency Detection Dataset**

The dataset used for filler word segmentation is obtained from Switchboard transcription . We also used the Automanner transcription for additional data. This gives more generalization to our training samples since it contains recording from standard interfaces. To label disfluent silences we use a combination of a silence probability model and a disfluency detection model

. First, we locate the silences and segment each word pair from the dataset then according to the probability model it's decided if silence is disfluent. For each word pair utterance the silence probability model gives a probability of a silence (Psil) occurring between them. A word pair with low Psil but a significant amount of silence is labeled as disfluent. If a word pair doesn't exist in the model vocabulary, we resort to the following approach. Since, general disfluencies accompany longer silences, any silence within a disfluent segment is labeled as an unnatural pause. Additionally, the word pairs surrounded with silences more than 0.7 seconds are also labeled similarly. This choice is experimental and can be considered safe because it's considerably higher than the suggested quantitative measure of micro-pauses (fluent), 0.2 secs. On the other hand, additional fluent pairs are collected from TIMIT

- **Skills Extraction Dataset**

In this model we have built our own dataset by using Python's Selenium Web Scraper to collect a set of 6145 unique skills from different online courses websites such as (Coursera, Udacity, Lynda, & Udemy), we also went through hiring websites such as (Craigslist, Wuzzuf & Yellow Page) to collect descriptions of needed open job vacancies and collected a set of 200,000 job descriptions.

Examples:

- *Math teachers actively instruct students, create lesson plans, assign and correct homework, manage students in the classroom, communicate with parents, and help students prepare for standardized testing.*
- *We are looking for an ambitious sales professional who wants to grow and learn while making money. Always play station and sometimes counter strike, ping-pong, chess, risk, & monopoly.unlimited paid vacations.friendly management team.great team spirit and working environment.annual fun trips strategy workshops where full team*

participating in deciding where the company should go.a unique chance to learn how to build your startup company

6.3 Results

6.3.1 Emotion Recognition Model:

The model successfully returns emotions of each frame with accuracy of 71% and after adding all the results to a list, normalizing & scoring them we reached an accurate score representation of the total emotional status that ranges between 0 and 1.

6.3.2 Disfluency Detection Model:

The model successfully extracts disfluencies from an audio file version of the interview with the following performance

Features	Precision	Recall	F1
mfcc	0.9482	0.9610	0.9534

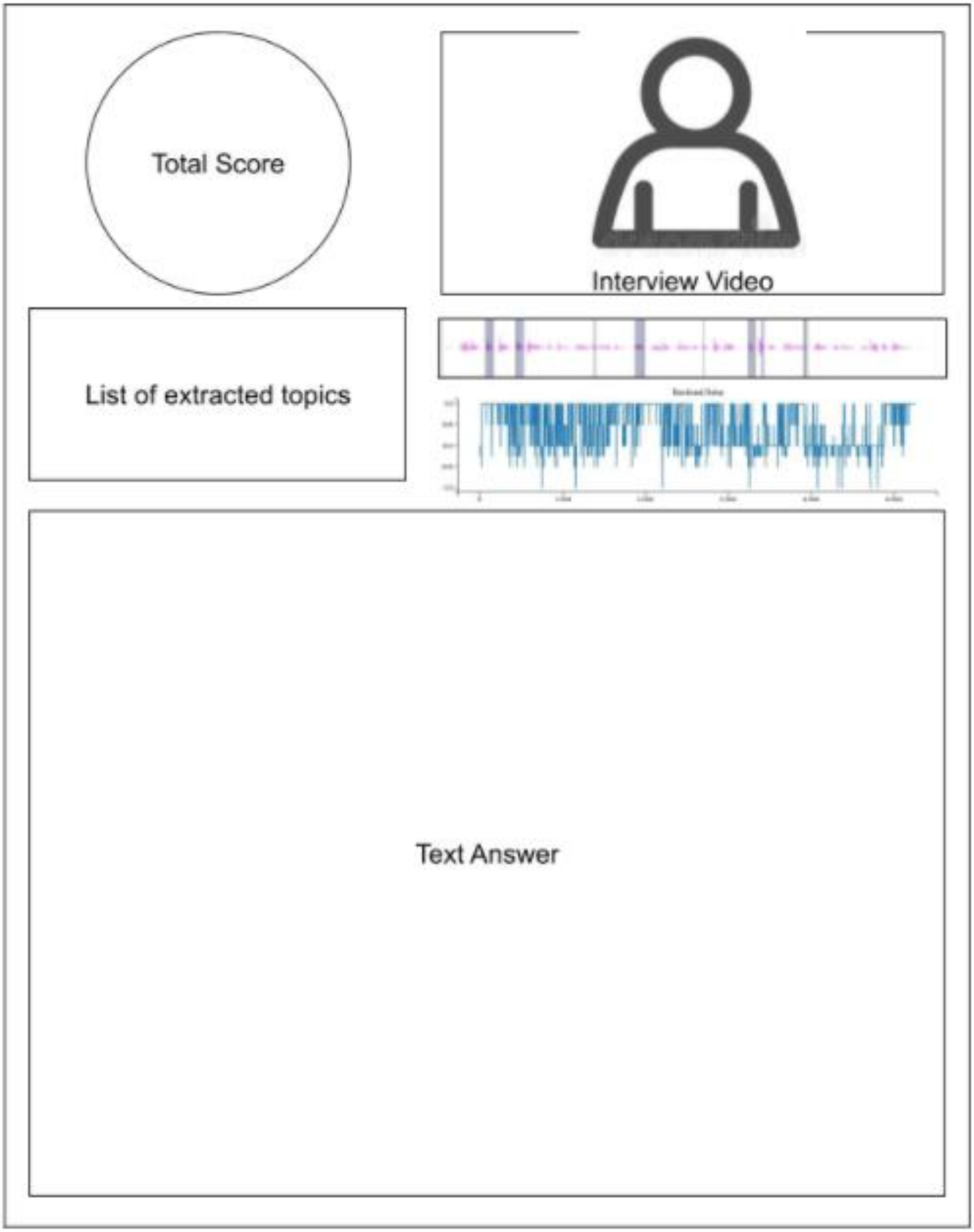
And we have calculated the score of the Disfluencies based on their count and their duration during the answer.
We have reached an accurate representation for the score that is ranged between 0 and 1

6.3.3 Skills Extraction Model:

The model was able to extract skills from a paragraph based on its context with an accuracy of 99%

6.3.4 The Report Page

The final output of our project is the report page where the final score of the models, the video of the interview, the text version of the interview,, the graph representations for both the disfluencies and the emotional status and finally the skills or topics mentioned in the answer are displayed in a way designed for maximum efficiency for the eyes for the hiring manager.
where they can quickly and accurately scrim through the report and analyze the answer in a matter of seconds. which was the main goal of our project that we have successfully achieved


Total Score
Interview Video
List of extracted topics
Text Answer

Chapter 7

Conclusion and Future work

This chapter concludes the overall benefits of the platform and conducts research for the future work.

The structure of this chapter is as follows: **Section 1** explains the conclusion and talks about the benefits of IVI. **Section 2** suggest future work that will enhance the performance of the platform

7.1 Conclusion

IVI is a smart interviews platform that can do analysis of interview videos and outputs a report that helps Employers pick the most fitting candidates.

Having a website like that helps doing such tasks can greatly reduce the effort and time needed to filter a huge amount of candidates for the next recruiting phase.

The output report contains the video of the interview, the speech to text version, Graph representing the emotional status, graph representing the disfluency intervals, skills extracted from the interview and an overall score for the answer.

Using deep learning algorithms such as CRNN & CNN & LSTMS we have reached an accuracy of 71% on the emotional status model, 99% validation accuracy on the skills extraction model & 0.9534 F1 measure score for the Disfluency Detection model.

7.2 Future Work

Our platform performs with a decent accuracy and overall performance, along the way we discovered some tweaks that would enhance the platform performance

- Improving the emotional status model by focusing more on micro emotion and focus on the change of facial features on small time periods to help capture more of the change of moods after reading the question to detect the element of surprise or confidence.
- Using more methods that can help in lie detection such as eye pupil movement.
- Currently our disfluency detection just detects Aahhs and Umms, in future versions the disfluency detection pipeline should also detect filler words such as “And”, “Okay”, “Right”, “like”... Etc. as well as silence fillers, word repeating fillers, giggles, and stuttering.
- Summarizing the text version of the interview and focusing on the key phrases in the answer.
- Detecting the accent of the candidate and how heavy it is.
- Detecting grammatical mistakes and wrong use of vocabulary and overall language fluency.
- Using the data collected from the interviews done on the website to build a model that can predict how likely the candidate will be accepted for the job.
- Using the set of answers and questions and acceptance/rejection ratio to detect correct and incorrect answers.
- Scoring the look of the candidate (hair, clothes and background) to see if they were appropriate for the interview or not.

- Using firebaseDB to make changes instantly appear on the Employers dashboard.
- Improving the scoring systems for models individually and the overall scoring system for the entire interview.

Reference

What does Emotion Recognition mean?	https://www.techopedia.com/definition/30819/emotion-recognition
Real-time Convolutional Neural Networks for Emotion and Gender Classification	https://arxiv.org/pdf/1710.07557.pdf
Blinking During and After Lying	https://link.springer.com/article/10.1007/s10919-008-0051-0
Eye blink detection with OpenCV, Python, and dlib	https://www.pyimagesearch.com/2017/04/24/eye-blink-detection-opencv-python-dlib/
HireVue: Pre-employment Testing & Video Interviewing Platform	https://www.hirevue.com
The Lips Don't Lie	https://www.psychologytoday.com/za/blog/spycatcher/200911/the-lips-dont-lie
Disfluent Responses to Job Interview Questions and What They Entail	https://www.tandfonline.com/doi/abs/10.1080/0163853X.2016.1150769?journalCode=hdsp20

Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection	https://arxiv.org/abs/1702.06286
Deep learning for specific information extraction from unstructured texts	https://towardsdatascience.com/deep-learning-for-specific-information-extraction-from-unstructured-texts-12c5b9dceada
A cognitive load approach to lie detection	https://onlinelibrary.wiley.com/doi/pdf/10.1002/jip.82
To Catch a Liar: Challenges for Research in Lie Detection Training.	http://www.communicationcache.com/uploads/1/0/8/8/10887248/to_catch_a_liar-_challenges_for_research_in_lie_detection_training.pdf
A Review of Emotion Recognition Using Physiological Signals	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6069143/
Survey on Emotional Body Gesture Recognition	https://arxiv.org/pdf/1801.07481.pdf
Increase Apparent Public Speech Fluency by Speech Augmentation	https://arxiv.org/pdf/1812.03415.pdf
Switchboard	https://www.iIVI.piconepress.com/projects/switchboard/w
TED Talk Analysis	https://roc-hci.com/current-projects/ted-talk-analysis/