Hiring Decision Analysis based on Interview Evaluation

HSN-620

Hemant Bidasaria Shubhanshu Shukla Vinay Joshi

Department of Humanities and Social Sciences Indian Institute of Technology, Roorkee

Multi Dimensional Hiring Decision Analysis Model

What is the Problem

Traditional hiring methods rely heavily on resumes, technical interviews, and gut feelings — often missing critical insights into a candidate's emotional intelligence, communication style, and real-time interaction behavior. This leads to:

- Poor cultural fit
- ☐ Weak communication alignment
 - Underperformance despite technical competence

Solution...

Comprehensive hiring model integrating:

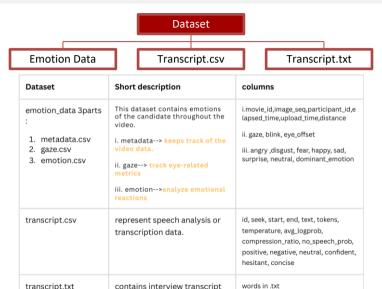
- Emotional Analysis (via facial expressions, gaze tracking, and emotional fluctuations)
- Communication Analysis (including speech pace, confidence, eye contact, hesitation, and conciseness)
- ☐ Transcript Analysis (using NLP to evaluate role-fit, technical depth, emotional tone, and language clarity)

Soft skills are increasingly more valued: According to LinkedIn's 2024 Global Talent Trends, 92% of HR leaders believe soft skills matter as much or more than hard skills **Team compatibility**: Matching communication and emotional patterns with company culture improves long-term retention.

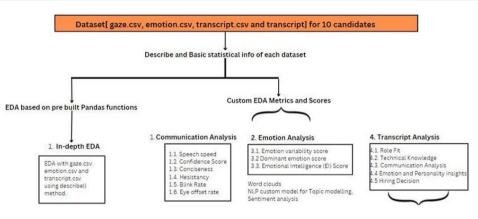
Recent Companies: HireVue, Pymetrics, Entelo, Retorio

27.35

Dataset used to create a multimodal model

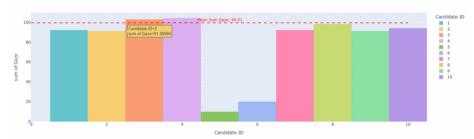


Flowchart of the model:



- ☐ A basic exploratory data analysis (EDA) was conducted to understand key metrics like mean, count, standard deviation, and percentiles for each candidate's interview data.
- Scores for all other parameters are calculated considering various parameters as will be discussed further.

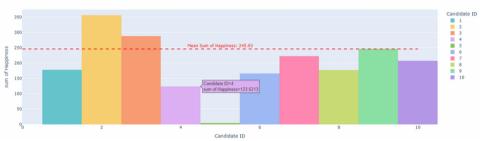
Mean of Gaze Sum for each candidate: (EDA on Emotion combined.csv)



- ☐ The plot shows the total sum of gaze, indicating how often each candidate looked at the camera during an interview.
- Candidates 5 and 6 have notably lower gaze values compared to others. This suggests that these candidates might not have been focused on the camera, potentially reading from a device or paper.
- ☐ Possible External Factors: Candidates with lower gaze values might have experienced distractions, technical difficulties, or external factors affecting their eye contact, which could be worth investigating.

Plot of Happiness distribution among candidates:

Happiness Distribution Across Candidates

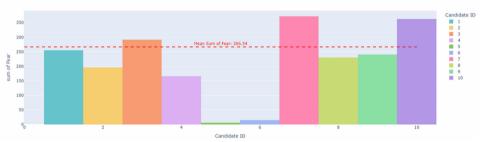


- The plot suggests that candidates with happiness above the average line should be considered for roles that benefit from higher emotional engagement.
- High levels of happiness suggests they could potentially be better suited for roles that require a positive disposition or team collaboration.
- Below-Average Candidates could indicate areas for concern if happiness is an important trait for the role in question.

6/35

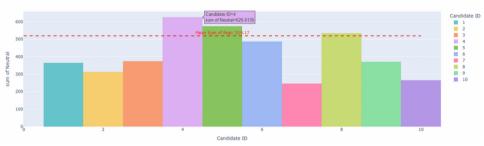
Plot of Fear distribution among candidates:

Fear Distribution Across Candidates



- ☐ This plot shows the fear distribution across di erent candidates, with each bar representing the sum of fear detected for each candidate
- ☐ Depending on the nature of the job, moderate levels of fear could indicate a healthy awareness of challenges, while extreme levels (either high or low) might require more investigation into how the candidate deals with pressure.
- Above-Average Fear Levels (Candidates 7, 10, 3, and 1): If fear is associated with stress or anxiety in a work context, these candidates may require further exploration to understand how they manage fear or stressful situations

Plot of Neutral distribution among candidates:



- ☐ High Neutrality: Candidates 4, 5, and 8, with higher neutrality, might be ideal for roles requiring calmness under stress, impartial decision-making, or situations where maintaining neutrality is crucial.
- Low Neutrality: Candidates who exhibit lower levels of neutrality may either be more passionate or emotionally expressive, which can be beneficial in roles that require strong emotional engagement (such as leadership or creative roles).

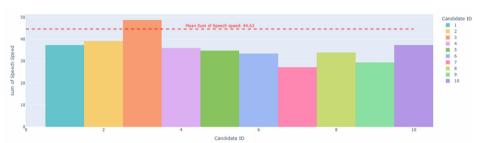
Creating Scores for candidates

Overall Candidate Scores (Based on emotional stability, fear handling, and happiness):

happiness):										
Candidate	Happiness Score (Out of 5)	Sadness Score (Out of 5)	Fear Score (Out of 5)	Neutrality Score (Out of 5)	Overall Score					
Candidate 1										
Candidate 2					3.5					
Candidate 3					3.5					
Candidate 4					4.5					
Candidate 5					4.5					
Candidate 6					4					
Candidate 7					3.5					
Candidate 8					3.75					
Candidate 9					4.25					
Candidate										

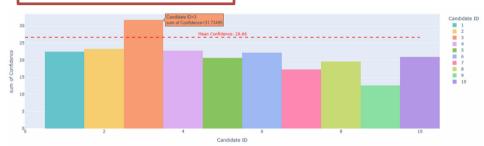
Based on the job profile we can give weights to different emotional scores and calculate overall candidate score impacting its hiring.

Speech speed distribution across candidates: (EDA on Transcript combined.csv)



- ☐ Above-Average Speech Speed: Excessively fast speech can sometimes be difficult to follow, particularly in roles that require clear communication.
- Below-Average Speech Speed: While slow speech can be perceived as a sign of calmness and careful consideration, if it's too slow, it may also suggest hesitation or lack of confidence, depending on the context.
- ☐ Balanced Speech Speed: Candidates like 4, 5, and 6 are closer to the average, indicating a more neutral or balanced pace. This might be desirable in roles that require clear communication and an ability to articulate thoughts at a reasonable speed.

Confidence distribution across candidates:



- □ Above-Average Confidence: These candidates might be suitable for roles requiring leadership, decision-making, or client-facing responsibilities.
- Below-Average Confidence: These candidates might be more reserved or unsure in their responses, which may suggest they are less comfortable under pressure or need more development in assertiveness for roles that demand confident communication.
- ☐ Similarly for hesitant emotion, conciseness emotion, etc we can do the analysis

Creating Scores for candidates

Overall Candidate Scores (Based on all four communication metrics):									
Candidate	Speech Speed Score (Out of 5)	Confidence Score (Out of 5)	Hesitation Score (Out of 5)	Conciseness Score (Out of 5)	Overall Score				
Candidate 1	4				3.25				
Candidate 2					4.25				
Candidate 3	4				4.25				
Candidate 4					4.25				
Candidate 5					4				
Candidate 6					4.25				
Candidate 7					3.5				
Candidate 8					3.5				
Candidate 9					3.5				
Candidate 10					4				

Based on the job profile we can give weights to different communication metrics and calculate overall candidate score impacting its hiring.

12/35

Why?

Goal: systematically extract and quantify key insights from candidate interviews:

- Sentiment
- Communication skills
- Topic relevance
- · Technical competencies

Method: using natural language processing (NLP) + AI tools.

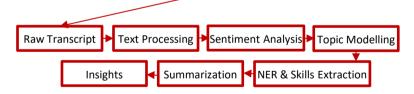
→ Enables objective, data-driven, and efficient evaluation of candidates, helping organizations identify top talent, ensure fair and unbiased assessments, and make well-informed hiring decisions that align with role requirements and organizational goals

Why This Matters:

- Reduces bias by converting subjective interviews to quantifiable metrics.
- 63% faster candidate screening vs. manual methods (LinkedIn's 2024 Global Talent Trends)

Transcript Analysis Workflow for Hiring Decisions

Hello, I am Jeffrey Shepherd and I am currently pursuing postgraduate and management from IIM Coikode. I have completed my B.Tech in Biotechnology from Heritage Institute of Technology Kolkata, followed by my M.Tech from IIT Kharagpur. I come with an experience of three years in the regulatory affairs domain of the pharmaceutical industry and I worked as a medical writer in Ciro Klein Farm, Mumbai and have specialized in drug safety and risk management. What sets me apart is the expertise I bring in with my three years of experience and an added two years of postgraduation. Along with this, I add another dimension to the discussion with my background



Transcript Analysis Workflow for Hiring Decisions



1. Text Processing

- Tokenization: Splits text into words/sentences (word_tokenize)
- ☐ Stopword Removal: Filters out 120+ non-essential words (e.g., "the", "is")
- $lue{}$ Lemmatization: Converts "running" \Rightarrow "run" using SpaCy's linguistic database

Why It Matters:

☐ Reduces data noise by 40-60%, and standardizes responses for accurate NLP model input

Candidate 1 Output:

- ['hello', 'jeffrey', 'pursue', 'management', 'iim', ...]
 - ☐ Interpretation: Cleaned data reveals focus areas like "management" and academic background ("IIM")

15/35

Transcript Analysis Workflow for Hiring Decisions



Why VADER?

- Optimized for conversational text (social media/transcripts)
- ☐ Captures intensity via capitalization/punctuation ("GREAT!!" \rightarrow +0.8 vs "good." \rightarrow +0.3)
- ☐ Processes 10K+ words/sec vs ML models (critical for bulk interviews)

How it works and Key Metrics:

from vaderSentiment import SentimentIntensityAnalyzer
analyzer = SentimentIntensityAnalyzer()
sentiment = analyzer.polarity_scores(transcript)
Output: {'neg': 0.825, 'neu': 0.889, 'pos': 0.167, 'compound': 0.9914}

Metric	Range	Candidate 1 Score	Interpretation
Compound	[-1,1]	0.9914	Strong positivity
Positivity Ratio	%	16.7%	Moderate enthusiasm
Neutrality	%	80.9%	Balanced responses

16/35

Transcript Analysis Workflow for Hiring Decisions



Interpretation Framework

- \square High compound (>0.5): Confident speakers (e.g., Candidate 1: 0.99 \rightarrow Pharma client roles)
- Low compound (<-0.1): Flag for HR review (potential hesitation/uncertainty)</p>
- Neutral dominance (>70%): Technical roles over sales

Why not others?

Handles interview-specific constructs: "Not bad" \rightarrow +0.35 (TextBlob: -0.3)

Tool	Limitation	Hiring Context Impact	Tool
TextBlob	Misses modifiers ("extremely good")	Underestimates candidate enthusiasm	TextBlob
BERT/LLMs	Slow (3 mins/transcript)	Impractical for 1000+ applicants	BERT/LLMs
Custom ML	Requires labeled interview data	High setup cost for SMEs	Custom ML

Transcript Analysis Workflow for Hiring Decisions



What is Topic Modeling?

- ☐ Topic modeling is an **unsupervised machine learning** technique that automatically discovers the main themes or "topics" present within a large collection of unstructured text, such as interview transcripts.
- ☐ Each topic is represented by a cluster of keywords that frequently co-occur, helping to summarize and interpret the content of documents at scale

Why Use Topic Modeling in Hiring Analysis?

- · Extracts Key Themes
- Objective Comparison
- Role Fit Assessment
- Scalable Insight

Transcript Analysis Workflow for Hiring Decisions



How it works?

- Document-Term Matrix: A matrix is created to represent word frequencies across all transcripts.
- Model Fitting: LDA is applied. LDA assumes each document (transcript) is a mixture of topics, and each topic is a distribution over words.
- ☐ Topic Extraction: The model outputs topics as sets of keywords, along with the proportion of each topic present in each transcript

Transcript Analysis Workflow for Hiring Decisions



What is LDA (Latent Dirichlet Allocation)?

- ☐ Initialization: Decide on the number of topics you want the model to find (e.g., 3–5 topics). Randomly assign each word in each document to one of the topics.
- ☐ Iterative Topic Assignment: For each word in each document:

Reassign the word to a topic based on: How often the word appears in each topic across all documents. How often topics appear in the current document. Repeat this process for many iterations so that word-topic assignments stabilize.

- ☐ Topic Formation: After enough iterations, each topic becomes a distribution of words (keywords with high probability in that topic). Each document becomes a mixture of topics (e.g., 60% Topic A, 30% Topic B, 10% Topic C).
- Output: For each topic, a List of top keywords that define it. For each document:
 Proportion of each topic present.

Transcript Analysis Workflow for Hiring Decisions



Example Output (from Candidate 1):

- ☐ Topic 1 (18%): "experience", "work", "management", "detail"
- Topic 2 (13%): "professional", "career", "attention"
- Topic 3 (9%): "biotechnology", "research", "kharagpur"

Interpretation: This candidate's transcript focuses on work experience, attention to detail, and technical research, indicating a strong fit for technical or management roles.

21 / 35

Transcript Analysis Workflow for Hiring Decisions



- Named Entity
 Recognition:
 Identified key
 entities such as
 technologies,
 companies, and skills
 from the text.
- Skill Extraction:
 Extracted relevant
 skills from the
 transcripts using
 keyword matching
 and NLP techniques.

Candidate	Sentiment	Skills	Verdict
Candidate 1	Neutral with slight positivity	Focus on work experience and growth	Suitable for mid-level roles in management or operations, strong experience but lacks enthusiasm.
Candidate 2	Slightly positive	Focus on internships, startups, and business acumen	Suitable for entry-level or junior roles in startups, strong in internships but lacks depth.
Candidate 3	Neutral to positive	Adaptability, problem-solving, personal/professional growth	Suitable for dynamic roles requiring adaptability and continuous learning.
Candidate 4	Neutral with slight positivity	Strong in analysis, data, and learning	Suitable for data analysis or research roles, strong analytical and learning focus.
Candidate 5	Positive	Focus on innovation and problem-solving, particularly with	Suitable for creative or innovative tech roles requiring problem-solving.

ΑI

Transcript Analysis Workflow for Hiring Decisions



Candidate 4: strong focus on their academic background, work experience, and project involvement. They appear to have a technical or engineeringrelated profile. This profile suggests that the candidate might be wellsuited for technical roles or positions requiring analytical and project management skills.

Word Cloud for Candidate 4



Feedforwarding results to AI

Here are the scores for each candidate based on the provided analysis, with a score scale from 1 to 5 (5 being the best) across the key metrics of Sentiment, LDA Topics, NER Entities, Skills, and Final Role Fit.

Candidate	Sentiment Score	LDA Topics Score	NER Entities Score	Skills Score	Final Role Fit Score
Candidate 1		4	4	4	3.5
Candidate 2	4	4.5	4.5	4	4
Candidate 3	4	4	4	4.5	4
Candidate 4	4				4.5
Candidate 5	4.5	4	4.5		4.5
Candidate 6	4				4.5
Candidate 7	3.5	4	4	4	4
Candidate 8	4	4	4.5	4	4
Candidate 9	4.5				4.5
Candidate 10	3.5	4	4	4	4

Objective:

To measure communication effectiveness using behavioral indicators like confidence, speech, eye movement, and hesitation.

```
•Scoring Method:

combined_conf_score = w1*avg_confidence_score -

w2*(std_confidence_score/avg_confidence_score)
```

data.head()

d	seek	start	end	text	tokens	temperatu	avg_logpr	compress	no_speecl	positive	negative	neutral	confident	hesitant	concise	enthusias	speech_speed
0	0	0	4.84	Hello, I'm	[50364, 24	0	-0.27553	1.583333	0.169504	0.539392	0.143638	0.31697	0.897895	0.924957	0.740309	0.693109	2.892562
1	0	4.84	8.96	with spec	[50606, 36	0	-0.27553	1.583333	0.169504	0.82568	0.104166	0.070154	0.732468	0.298984	0.556738	0.388407	1.941748
2	0	8.96	12.88	certificati	[50812, 21	0	-0.27553	1.583333	0.169504	0.685738	0.156997	0.157265	0.718465	0.414894	0.471627	0.421808	2.295918
3	0	12.88	16.88	second o	[51008, 11	0	-0.27553	1.583333	0.169504	0.464676	0.354674	0.18065	0.1339	0.584103	0.340603	0.08333	3.5
4	0	16.88	22	busy doin	[51208, 58	0	-0.27553	1.583333	0.169504	0.423742	0.27622	0.300038	0.837648	0.58261	0.352158	0.311778	2.34375
5	0	22	26.88	various th	[51464, 36	0	-0.27553	1.583333	0.169504	0.643381	0.100235	0.256385	0.640092	0.369793	0.046411	0.45998	2.254098
6	2688	26.88	30.84	personal	[50364, 29	0	-0.29564	1.667732	0.45836	0.686997	0.100657	0.212346	0.733875	0.662218	0.318639	0.464803	2.777778
7	2688	30.84	34.28	internatio	[50562, 50	0	-0.29564	1.667732	0.45836	0.686823	0.134788	0.178389	0.64676	0.53998	0.215795	0.41216	3.197674
8	2688	34.28	38.48	called Kre	[50734, 12	0	-0.29564	1.667732	0.45836	0.611754	0.164233	0.224013	0.778986	0.31412	0.335227	0.671075	3.333333
9	2688	38.48	42.44	biggest pr	[50944, 38	0	-0.29564	1.667732	0.45836	0.950391	0.013648	0.03596	0.881679	0.007717	0.377374	0.903439	2.272727
10	2688	42.44	47.5	awarenes	[51142, 88	0	-0.29564	1.667732	0.45836	0.584468	0.286314	0.129218	0.197129	0.73429	0.245009	0.128551	2.173913
11	2688	47.5	51.76	idea to sh	[51395, 15	0	-0.29564	1.667732	0.45836	0.671625	0.084068	0.244307	0.65023	0.301754	0.405086	0.591487	4.225352
12	2688	51.76	55.44	and what	[51608, 29	0	-0.29564	1.667732	0.45836	0.10482	0.755571	0.139609	0.065768	0.631814	0.214568	0.011457	2.98913
13	5544	55.44	59.04	and then	[50364, 29	0	-0.34832	1.571429	0.004393	0.666163	0.150201	0.183636	0.417246	0.592568	0.297613	0.227758	4.166667
14	5544	59.04	63.08	students	[50544, 17	0	-0.34832	1.571429	0.004393	0.537353	0.267005	0.195642	0.62511	0.685129	0.344343	0.187111	2.970297
15	5544	63.08	67.52	teaching:	[50746, 45	0	-0.34832	1.571429	0.004393	0.937737	0.011025	0.051238	0.737882	0.184195	0.809419	0.460207	2.702703
16	5544	67.52	74.52	activities	[50968, 53	0	-0.34832	1.571429	0.004393	0.698996	0.075779	0.225224	0.336462	0.016167	0.961038	0.011409	1.857143

Key Metrics:

- Confidence Score: Measures consistency of confidence in speech
- Blink Rate: Should be moderate too high signals nervousness
- Eye Offset Score: Consistent eye alignment indicates focus
- Speech Speed: Should be neither too fast nor too slow
- Conciseness: Encourages brief and meaningful communication
- Hesitancy: Moderate pauses imply thoughtful, structured speech

Interpretation:

- High average values → Good performance
- Low variability → Consistency and stability
- Penalizes erratic or inconsistent behavior

Confidence Score Average:

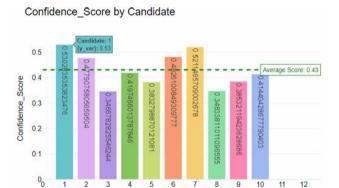
•Generally, a confidence score above 0.7 suggests that the candidate is performing well and making predictions with reasonable certainty.

Confidence Score Variance:

•A low variance (like 0.0434) indicates that the confidence scores are relatively consistent. This means the model is not fluctuating wildly in how confident it is in its predictions, which is a good sign of stability

28 / 35

Hiring Decision Analysis HSN-620 April, 2025



The combined confidence scores reflect communication e ectiveness. Candidates 3, 5, 8, and 9 show lower confidence levels.

Candidate

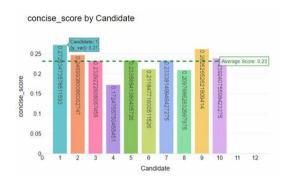
Hiring Decision Analysis HSN-620 April, 2025 29/35

 $combined_speech_score=w1*avg_speech_speed-w2*std_speech_speed$



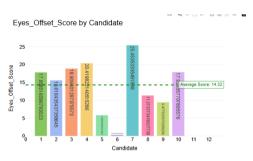
Candidates 3, 2, and 10 should be considered for roles that require strong communication, as they performed well in this area. Conversely, Candidates 7 and 6 may need improvement in speech clarity and should be considered for roles where communication is less critical.

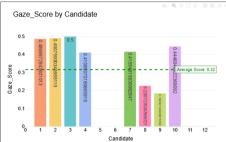
Concise comm---> a good communicater



Conciseness is a key metric for evaluating communication skills. However, candidates 4, 6, and 8 performed below others in this area.

Hiring Decision Analysis HSN-620 April, 2025 31/ 35





Eye offset indidcates the deviation the candidate is looking away from the camera and this along with the Graze score confirms that Candidate 7 has both below average graze score and is looking away from the camera

Emotional Intellegence Score

We assigned emotional weights to each emotion and summed them across the video to track emotional fluctuations and assess the candidate's emotional intelligence.

Approach to Calculate Combined Score:

- 1. Mean (µ): Represents the average emotional intelligence of the candidate.
 - Higher is better: Indicates positive or emotionally intelligent behavior.
- 2. Standard Deviation (a): Represents the variability or consistency in emotional intelligence.
 - Lower is better: Indicates more emotional stability and consistent behavior.

Combined Score Formula:

To combine these two factors, the mean is the main indicator of emotional intelligence, and the standard deviation is a penalty for inconsistency. We can calculate the combined score as:

Combined Score =
$$\mu - k \times \sigma$$

Where:

- µ is the mean El score.
- σ is the standard deviation.
- k is a penalty factor (a constant to control how much weight you give to inconsistency).

Emotional Intellegence Score

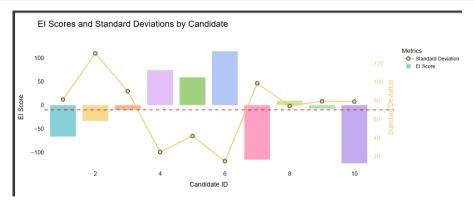
```
def calculate_ei(row):
    # Assign weights to emotions
    emotion_weights = {
        'angry': -2, # Negative influence
        'disgust': -2, # Negative influence
        'fear': -1, # Negative but less intense
        'happy': 2, # Positive
        'sad': -2, # Strong negative
        'surprise': 0, # Neutral
        'neutral': 1 # Moderate positive
```

Calculate score based on emotions and intensity emotional_score = sum(row[emotion] * weight for emotion, weight in emotion_weights.items()

Emotional weights were assigned to each emotion and they were cumulatively summed over the entire duration of the video. This made us understand how emotions of the candidate fluctuated

34/35

Emotional Intellegence Score



El Score (bar chart): Reflects each candidate's emotional intelligence based on their emotional perception, regulation, expression, and dominant emotions during the interview. Standard Deviation (line with dots): Shows the volatility or consistency in a candidate's emotional intelligence across the interview. A lower standard deviation indicates emotional stability, while a higher standard deviation points to emotional volatility or inconsistency

Conclusion/ Results

The final decision of comprehensive hiring would include scores from all the steps:

Prebuilt in Function EDA Communication Analysis Score

Emotional Intelligence Score Transcript Analysis Score

- ☐ The scores obtained in each of the metrics are very much dependent upon the weights we put in each subpart, that depends on how much that metrics essence is needed in the job role for which the hiring is being made.
- ☐ Also, a cumulative score combing the four above metrics is also based on role, thereby creating a formula to get the final score according to the role fit.
- ☐ Instead of keeping a threshold based on value we can keep it on the basis of requirement for the company and industry standards.

Conclusion/ Results

	Gaze Sum /	Emotional	Communication	EI	Overall	Average
Candidate	Eye Offset	Stability	Skills	Score	Role Fit	Score
Candidate 1	3.5		3.25	3.5		3.05
Candidate 2	4	3.5	4.25	4.5	4.9	4.23
Candidate 3	4	3.5	4.25		2.4	3.23
Candidate 4	4.5	4.5	4.25	4.5	4.2	4.39
Candidate 5	2.5	4.5	4.25	4.5	4.7	4.09
Candidate 6	3.5	4	4.25	1.5	4.3	3.51
Candidate 7	3.5	3.5	3.5	2.5	4.3	3.46
Candidate 8	4	3.75	3.5	1.5	4.9	3.53
Candidate 9	4.5	4.25	3.5	4.5	4.1	4.17
Candidate 10	3.5		4		4.1	3.32

Thank You!! We are open to questions.