

Basic Details

Candidate Name: ***** ****

Panel: ADA

Candidate Email: *****@gmail.com

Position: Data Scientist (NLP) Ujjwal

Candidate Phone: +91-*****

Time of Interview: 6:00 PM

Date of Interview: 27-04-2023

Interview Status: Strong Hire

Overall Rating: 4.36 / 5

Interview link: <https://futuremug.co.in/demo/demo1.mp4>

Scoring



Behavioral

Communication



The candidate is a little slow in communication but this should not be a deal breaker

Industry Awareness



The candidate has very good knowledge of the industry and the leading techniques used

Engineering Mindset



The candidate has a very good mindset for tackling engineering problems.

Attitude



The candidate was able to demonstrate a good attitude during the interview process

Team Work

The candidate has worked on multiple projects in teams before and has the relevant experience



Problem Solving

During the coding round, the candidate was able to very quickly understand the given problem and solve it



Analytical Skill

The candidate was able to showcase very good analytical thinking while discussing various NLP topics



Technical Skill Set - Mandatory

C++

The candidate has multiple years on experience in C++



mongodb

The candidate has previously worked with mongoDB in multiple projects



Machine Learning

The candidate has previously worked with ML techniques in multiple projects



NLP

The candidate's core strength is NLP as he has worked on many projects in the past.



Mysql

The candidate was able to convey that he as the relevant experience of working on MySql



GCP

The candidate has no prior experience working in GCP



bash

The candidate has very limited experience working in bash



Pytorch

The candidate has worked on many projects that involved working on Pytorch



keras

The candidate has worked on many projects that involved working on KERAS



Pandas

The candidate has worked on many projects that involved working on PANDAS



Numpy

The candidate has worked on many projects that involved working on NUMPY



Tensorflow

The candidate has worked on many projects that involved working on TENSORFLOW



BERT

The candidate has worked on many projects that involved working on BERT



precision

The candidate is aware on how precision works in ML optimization techniques



recall

The candidate is aware of how recall works in ML optimization techniques



WER

The candidate has previously worked on WER and has worked on many similar projects.



BLEU score

The candidate has very limited experience in this field



ROUGE

The candidate has very limited experience in this field



METEOR



The candidate has very limited experience in this field

AUC-ROC



The candidate is aware of how AUC-ROC works in ML optimization techniques

F1-score



The candidate is aware of how F1 score works in ML optimization techniques

coherence



The candidate is aware of how the coherence technique works in ML optimization

plotly



The candidate has worked on many projects that involved working on plotly

Matplotlib



The candidate has worked on many projects that involved working on matplotlib

Question & Answers

Question

Candidate Answer

How are Word2Vec embeddings different than GloVe embeddings?

Word2vec embeddings are based on training a shallow feedforward neural network while glove embeddings are learnt based on matrix factorization techniques

Correct Answer

What is the difference between having an auto-encoder or auto-regressive models?

The only difference is how the model is pre-trained. The same architecture can be used for both the models.

Correct Answer

What are the advantages of using BiLSTM over LSTMs?

In Bidirectional LSTM we can input the data in both the direction and hence it is more powerful than LSTM.

Correct Answer

Was IR used in knowledge graph, what algorithms were used?

A knowledge graph is a representation of entities (e.g., a person, a place) and the relationships between them. Knowledge graphs are used to derive semantic understanding from these connections.

Correct Answer

What prompting techniques were used in LLMs?

Correct Answer

Prompts are the inputs or queries that a user or a program gives to an LLM AI, in order to get a specific response from the model. Prompts can be NL sentences or questions, code snippets/commands, combination of text or code, depending on the domain and the task.

Naive Bayes implementation?

Correct Answer

It is used in sentiment analysis

What are counterfactual generative models?

Correct Answer

Couldn't clearly tell about the model, ended up explaining conditional generative model

Basic architecture of CNNs

Correct Answer

It has three layers namely, convolutional, pooling, and a fully connected layer. CNN have grid like typology.

What are the differences between RNNs, GRUs and LSTMs

Correct Answer

The key difference between RNNs, LSTMs, and GRUs is the way that they handle memory and dependencies between time steps. RNNs remember information from previous inputs but will become very slow with long-term dependencies. LSTMs can store and deal with long-term dependencies using a special type of memory cell and gates. GRUs is simplified version of LSTMs and uses single update gate and is therefore easier to train and run.

What are the differences between tokenizers from nltk and huggingface tokenizers?

Correct Answer

Candidate has little information on the difference between them. He was able to explain how tokenisation works.

GRU and LSTMs difference in architecture

[similar question] Already discussed above.

Correct Answer

GRU and LSTMs difference in architecture

[same question as above] Already discussed above.

Correct Answer

Transformers vs RNNs

Transformers also make use of attention mechanisms but RNN does not.

Correct Answer

What is self attention?

In a single sentence, how the words are related to each other. this is gathered by Self attention.

Correct Answer

Different ways to get word embeddings?

One-hot encoding, tf-idf, word2vec, fasttext

Correct Answer

Do vectors hold semantic meanings?

Candidate was able to answer YES. But could not go in detail when asked to take an example and explain.

Correct Answer

LLMs are said to have a lot of parameters, what does it mean?

The parameters are basically the variables present in the ML to be trained.

Correct Answer

What are various optimization algorithms to decide the Adam optimisation, learning rate decay, gradient

learning rate?

descent

Correct Answer

How is BERT different than GPT-3?

Gpt-3 is an autoregressive model, while BERT is bidirectional

Correct Answer

Coding Skills

```
print("hello, world")
```

```
string = ""
```

```
# Avatar
```

```
# character_frequency("abbcba") => "a2b3c1" character_frequency("aaabbb") => "a3b3"
```

```
def character_frequency(string):
```

```
    char_count = {}
```

```
    for char in string.lower():
```

```
        if char in char_count:
```

```
            char_count[char] += 1
```

```
        else:
```

```
            char_count[char] = 1
```

```
    output = ""
```

```
    for char, count in char_count.items():
```

```
        output = output + f"{char}{count}"
```

```
    return output
```

```
res1 = character_frequency("abbcba")
res2 = character_frequency("aaabbb")
print(res1)
print(res2)
```

```
#Avatar
```

```
"""If the term "apple" appears 50 times in a document containing 1000 words,
and it appears in 10 out of 100 documents in a corpus, what is the TF-IDF
score for "apple" in the given document or corpus? """
```

```
import numpy as np
tf = 50 / 1000
idf = np.log(10 / 100)
tfidf = tf * idf
```

```
print(f"tf: {tf}")
```

```
print(f"idf: {idf}")
```

```
print(f"tfidf: {tfidf}")
```

Output

```
hello, world a2b3c1 a3b3 Traceback (most recent call last): File "script.py", line 34, in import numpy as np
ModuleNotFoundError: No module named 'numpy'
```

Candidate Screenshot



Overall Comments

The candidate is very good and has a strong foundation in the fundamentals of NLP. Although the candidate was unable to answer some of the theoretical questions and also needs improvement in basic communication I still feel that he will be a perfect fit for the role as per the JD given. There could be a learning curve associated with his hiring, but he will be able to pick it up soon and become an independent contributor to the team. I still feel that given his previous experience, he should be the right resource for this job.