

LLM Project

Predict if a COVID patient requires ventilator

Medicare Issue

- We have seen during COVID 19 that plenty of patients died due to lack of ventilators. Doctors were overworked and could not attend to patients timely.
- The issue I try to solve in this project is to build a predictor that tells if a patient needs a ventilator or not based on observations and conditions. If we can build a model that can predict ventilator usage, it will help ease some stress off of doctors/nurses and also help hospitals plan for resources.

Dataset Used

- I used 10K COVID related synthesized data provided in the assignment instructions.
- From this dataset, I used below files
 - devices.csv: used to get ventilator target field
 - conditions.csv: used to get patient conditions like Acute Respiratory Failure
 - observations.csv: used this file to get patient observations like ferritin in serum or plasma, lactate dehydrogenase, oxygen in arterial blood and oxygen/inspired gas setting.

Prompts that are engineered

I engineered prompts to predict my target variable, i.e., if a patient requires a ventilator or not depending on the features like ferritin in serum or plasma, lactate dehydrogenase, oxygen in arterial blood and oxygen/inspired gas setting.

Prompts tried:

- First I tried below prompt without just user role `Decide in a single word if the patient requires a ventilator given the patient does not have Respiratory Failure. Ferritin in Serum or Plasma is 0. Lactate dehydrogenase is 0. Oxygen in Arterial blood is 0 and Oxygen/Inspired gas is 0. Does this patient requires a ventilator or not?`

This prompt returned wrong prediction. Next I used the same prompt with chain-of-thought.

- Then I tried chat based prompt and gave the system a role with content `You are a doctor and an expert on deciding if a patient requires ventilator.`

Even this approach lead to wrong prediction.

```
response = openai.ChatCompletion.create(  
    model="gpt-4",  
    messages=[  
        {"role": "system", "content": "You are a doctor and an expert on deciding if a patient requires ventilator."},  
        {"role": "user", "content": "Decide in a single word if the patient requires a ventilator given the patient does not have Respiratory Failure. Fe"}  
    ]  
)  
  
print(response.choices[0].message['content'])
```

No

Methods Employed: In-Context Learning

Chain-of-thought

Chain-of-thought: Since the prompt from last slide returned a wrong prediction, I tried chain-of-thought method with the same prompt to understand and better engineer my prompt.

```
messages = [
    {"role": "system", "content": "You are a doctor and an expert on deciding if a patient requires ventilator."},
    {"role": "user", "content": "'Decide in a single word if the patient requires a ventilator given the patient's Acute Respiratory Failure is 0. Ferritin in Serum or Plasma is 516.2. Lactate dehydrogenase is 257.1. Oxygen in Arterial blood is 37.6. Oxygen/Inspired gas setting is 64.2. Please decide whether this patient requires a ventilator or not.'"},
    {"role": "assistant", "content": response.choices[0].message['content']},
    {"role": "user", "content": "Can you provide details why the patient requires a ventilator?"},
]

response = openai.ChatCompletion.create(
    model="gpt-4",
    messages=messages
)

print(response.choices[0].message['content'])
```

➡ Based on the given information, the patient does not require a ventilator. A ventilator is typically used when a patient cannot breathe adequately on the
In this case, the patient has an Acute Respiratory Failure score of 0, indicating a lack of severe respiratory distress. The Ferritin and Lactate dehydro
The only measure somewhat concerning is the patient's Oxygen in Arterial blood which is at 37.6, which is low, as normal levels range from 75 to 100 mill
If the oxygen levels cannot be maintained despite supplemental oxygen, a ventilator might then be considered, but with the information currently provided
It's also worth noting that this decision should not be made based solely on these indicators but should involve a comprehensive review of the patient's

Based on the result from chain-of-thought, I changed my prompt to:

Decide in a single word if the patient requires a ventilator given the patient's Acute Respiratory Failure is 0. Ferritin in Serum or Plasma is 573.1. Lactate dehydrogenase is 312.5. Oxygen in Arterial blood is 51.0. Oxygen/Inspired gas setting is 83.8. Please decide whether this patient requires a ventilator or not.

This prompt helped get the correct prediction. And this also helped me change my input data. Originally I was passing 0 value for Oxygen in Arterial blood and 0 value for Oxygen/inspired gas setting if there was no observations for these. This was misleading as GPT was reading 0 as the value and suggested to use a ventilator. After I changed it to N/A instead of 0, the prediction was correct.

```
▶ response = openai.ChatCompletion.create(  
    model="gpt-4",  
    messages=[  
        {"role": "system", "content": "You are a doctor and an expert on deciding if a patient requires ventilator."},  
        {"role": "user", "content": "Decide in a single word if the patient requires a ventilator given the patient's Acute Respiratory Failure is 0. Fer  
    ]  
)  
  
print(response.choices[0].message['content'])
```

☞ Yes

Tree-of-thought

For this next approach I modified the prompt to Imagine three different experts are answering this question. All experts will write down 1 step of their thinking, then share it with the group. Then all experts will go on to the next step, etc. If any expert realises they're wrong at any point then they leave. The question is decide in a single word if the patient requires a ventilator given the patient's Acute Respiratory Failure is 0. Ferritin in Serum or Plasma is 1685.1. Lactate dehydrogenase is 313.3. Oxygen in Arterial blood is N/A. Oxygen/Inspired gas setting is N/A. Please decide whether this patient requires a ventilator or not.

```
# Tree of thought
messages = [
    {"role": "user", "content": '''Imagine three different experts are answering this question.
    All experts will write down 1 step of their thinking, then share it with the group.
    Then all experts will go on to the next step, etc. If any expert realises they're wrong at any point then they leave.
    The question is decide in a single word if the patient requires a ventilator given the patient's Acute Respiratory Failure is 0.
    Ferritin in Serum or Plasma is 1685.1. Lactate dehydrogenase is 313.3. Oxygen in Arterial blood is N/A.
    Oxygen/Inspired gas setting is N/A. Please decide whether this patient requires a ventilator or not.'''},
    {"role": "assistant", "content": response.choices[0].message['content']},
    {"role": "user", "content": "Can you provide details why the patient requires a ventilator?"},
]

response = openai.ChatCompletion.create(
    model="gpt-4",
    messages=messages
)

print(response.choices[0].message['content'])
```

Tree-of-thought Result

Expert 1:

Step 1: The fact that the patient has Acute Respiratory Failure severity 0 means that the patient is not currently experiencing severe respiratory distress, which is usually a primary indicator for the need of a ventilator.

Step 2: The serum Ferritin level is indeed very high, which can indicate inflammation or infection, but it is not typically used alone as a critical indicator for mechanical ventilation.

Step 3: The oxygen levels in arterial blood and inspired gas settings are unavailable. However given the ARF score and other values, and assuming that this score remained constant, it doesn't seem a ventilator would be required.

My decision would be: No

Expert 2:

Step 1: The elevated Ferritin level certainly could indicate a severe infection or other serious conditions.

Step 2: In the absence of additional symptoms or signs, particularly oxygen levels or clear signs of respiratory distress, it is however hard to justify the use of a ventilator solely on the basis of Ferritin and Lactate dehydrogenase levels.

Step 3: Even considering Ferritin and Lactate dehydrogenase levels, which may be associated with systemic inflammation or cellular damage respectively, there's no clear indication of lung function deterioration.

My decision would be: No

Expert 3:

Step 1: The severity of Acute Respiratory Failure (ARF) is 0, indicating the patient is not in severe distress.

Step 2: The Ferritin and Lactate dehydrogenase values, while exceeding normal range, may indicate inflammation or infection, but not necessarily reduced pulmonary function.

Step 3: Without values for arterial oxygen concentration or inspired gas fraction, it's difficult to make a definitive assessment of respiratory function, but the available information doesn't suggest ventilator need.

My decision would be: No.

It's important to note these steps represent individual analytical perspectives and the final decision should always be based on a comprehensive clinical assessment.

ChatGPT embeddings for Ventilator Prediction

Train test data split

- Did a random split of input dataset to use 80% of the data for training and 20% for testing.
- Made sure there are equal proportions of positive predictions in train set and test set.



```
data_index = list(df.index)
train_index, test_index = train_test_split(data_index, test_size=0.2, random_state=42)
```



```
print(train_index)
print(test_index)
```



```
[180, 17, 24, 176, 110, 73, 101, 137, 253, 271, 113, 218, 320, 245, 31, 268, 168, 148, 221, 22
[173, 132, 197, 9, 104, 119, 256, 158, 226, 312, 196, 90, 281, 84, 185, 269, 126, 78, 25, 5, 2
```

Get embeddings for train dataset

- Used “text-embedding-ada-002” model to generate embeddings for train dataset.
- Shape of the train dataset with embeddings is (258, 1536).
- Next we will use these embeddings as input to logistic regression.

```
[117] def generate_embeddings(texts, model="text-embedding-ada-002"):
      embeddings = []
      for text in tqdm(texts):
          text = text.replace("\n", " ")
          response = openai.Embedding.create(input = [text], model=model)['data']
          embeddings.append(response[0]['embedding'])
      return np.array(embeddings)
```

Get the embeddings of prompts of training dataset of endometrium

```
[119] train_indices = data_split['train indices']
      train_df = df.iloc[train_indices]
      train_ds = DrugCombDataset(train_df)

      embeddings = generate_embeddings(train_ds)

      100%|██████████| 258/258 [00:40<00:00, 6.39it/s]
```

Show the shape of the embeddings

```
np.shape(embeddings)
```

```
(258, 1536)
```

Logistic Regression

- Used logistic regression to train the model using embeddings generated from GPT.
- Testing this model on test data gives an ROC score of 96.9%.

Train a simple classifier to predict if ventilator is required or not using embeddings



```
from sklearn.linear_model import LogisticRegression  
  
model = LogisticRegression(max_iter=1000)  
model.fit(embeddings, labels)
```

```
▼ LogisticRegression  
LogisticRegression(max_iter=1000)
```

Test the performance

```
[ ] test_embeddings = generate_embeddings(test_ds)  
    test_labels = list(df.iloc[test_indices]['OUTPUT_LABEL'])  
  
    test_pred = model.predict_proba(test_embeddings)[:,-1]  
    auroc = roc_auc_score(test_labels, test_pred)  
    auprc = average_precision_score(test_labels, test_pred)  
    print('\nAUROC:', auroc, '\nAUPRC', auprc)
```

```
100%|██████████| 65/65 [00:09<00:00, 6.78it/s]  
AUROC: 0.9692982456140351  
AUPRC 0.9959577371127271
```