A3 - SAR Agent Modifications

Insights	Modifications
I had an incomplete requirements.txt, which made it difficult to install the required packages to run my code. Additionally, I did not inform the user or include a way to download Spacy's "en_core_web_sm" model.	I correctly updated requirements.txt, and added code to the agent's file to automatically download "en_core_web_sm" if it is not already found.
Gavin suggested that I return a list of predicted conditions with a confidence score attached to each of them. This allows for the user to see many possibilities of what they may be experiencing.	I used the predict_proba() method from scikit-learn's RandomForestClassifier to get the percentage of trees that voted for each condition. I then normalized the probabilities associated with the top 5 conditions to add up to 1.
The above suggestion gave me the idea to include the most common symptoms associated with the top conditions returned. This will give the user more information about what they may be experiencing, and can help them tailor their input.	In order to represent the top symptoms fairly, I decided to use a combination of frequent symptoms and unique symptoms for a given condition. I returned a list including the most frequent symptom, the symptom with the highest TF-IDF score, and the top 3 symptoms from the union of frequent list and the TF-IDF list.
In addition to top symptoms, I also wanted to include a description of the predicted condition. There was no dataset that had descriptions for all of the conditions, so I had to use wikipedia to get the best match.	I retrieved disease descriptions from Wikipedia by searching for the best-matching page title and extracting the summary. I stored the results in a CSV so they could be retrieved upon predicting a disease.
Lastly I wanted the output to be in a cleaner format, since my initial display was not very clear.	For this, I used a library called Rich. I split my output into 3 sections: left, middle, & right. The left side includes the top predicted disease, a description, and the top symptoms associated with it. The middle section contains a breakdown of the detected symptoms and how they were detected. The right section contains the top 5 predicted diseases with the associated probabilities.