Question:
You are asked to create a model (preferably deep learning, but traditional machine learning is also acceptable) to predict stroke hospitalization. Please provide explanations that justify your methods and results, and explain any assumptions made about the data. Present your findings in a clear and organized document. Include your code separately so that a reviewer can easily recreate the results.

Answer:
I use 2 machine learning models to predict stroke hospitalization. They are decision tree and MLP. I have 2 assumptions:

First, when I do the exploration of the data, my assumption 1 is that only clinical_event_time and clinical_event_member_age are useful for prediction. For other attributes, clinical_event_id and clinical_event_member_id are just index, member_male is all 1 which is all male, so these 3 attributes cannot make prediction.

Second, when I do further exploration and data modeling, my assumption 2 is that clinical_event_time is more important, because one clinical member may have several the same clinical_event_member_age, but have several different clinical_event_name, so only know clinical_event_member_age cannot tell which clinical_event_name for one member.

Question:
Explain why you choose the modeling approach and network architecture.

Answer:
This is just a simple supervised learning of classification, decision and MLP can both do classification, I can also use others like random forest, SVM and so on to do the same.

Question:
Report the accuracy of the model at predicting stroke hospitalization.

Answer:
We can see from our code that both two methods have 100% accuracy of prediction. This data is too simple, so that I get so high accuracy.

Question:
Describe which data attributes are predominately influencing stroke risk.

Answer:
We can see at the end of our code, clinical_event_time is predominately influencing stroke risk.