Problem
A significant fraction of patients who had heart failure is at risk of being readmitted to the hospital. A risk stratification model, i.e. identifying patients who are at higher risk of having another event, can be used by clinicians to identify high risk patients who need immediate treatment.

Task
1. Propose a solution on how can RWD (i.e. EHR data) be used to develop a risk stratification model.
- What kind of data would you use for this purpose? (You can assume EHR covers patient demographics, in-patient visits, diagnosis, prescriptions, lab measurements, & clinical notes)
- What is the modelling approach and how would you use the data in the model?

2. Notes or reports in the form of free text are commonly generated in clinical practice, which might contain information that could be useful in understanding patient condition and risk factors that are not captured in the structured database. Attached csv file contains samples of some radiology reports or notes.
- How would you use them to solve the problem above? What kind of information would you extract?
- Can you develop a proof-of-concept NLP pipeline that extracts information that you need for the model?

3. Can you propose the other ways RWD can be leveraged to reduce the disease burden? For example, can you suggest effective interventions to mitigate risk for high-risk patients?

At the end of one week, we expect you to share your code (github repository), a document and slides describing your approach and results, insights etc. The document and slides should follow a similar narrative. For this position, good visualization and communication skills are as important as analytical skills.

Best regards,
Lu

Lu Wang
Data Scientist | Holmusk
T: (SG) +65 8525 7236
E: lu.wang@holmusk.com

"Transforming lives through evidence-based medicine" | Follow us on LinkedIn.