Team Meeting

29 March 2023/ 2:15 PM / Melbourne Connect 3201

Attendees

Lida, Rohit, Nofaldi, Tony, Kaylee

Agenda

1. **Separate dataset with excel table ---- 30min**

- Manually split the original dataset into different datasets by first unpacking the A-AM column in the excel table into an Audit_ID corresponding to 39 features, remembering to anonymize the Audit_ID (no ID name appears, anonymize the ID as case1, case2 .....) The column names of these 39 FEATURES are first referenced to the column names in row 8 of the dataset, with the header in row 7 waiting for the next host meeting to ask for the meaning and the corresponding breakdown.
- Manually unseal the data from the AN to DB column into the data of the prescription point, you can use python's dataframe to unseal the data into 5 FEATURES corresponding to each AUDIT ID, there is a column name about the OLDER in the line of the prescription point, first temporarily IGNORE, wait for the next MEETING with the HOST and then ask. The remaining columns are divided into different tables by referring to step 2.
- Concatenating the above tables, for example the table for the prescription point will have 44 features corresponding to a label (prescription point) as the final clean data dataset, and then the dataset will be processed for null values. Write a readme file to explain the features and labels in this table.

2. **Data analysis after the clean dataset ---- 10min**

- The cleaned data is first sorted into features, then the distribution of the different features is analysed and plotted, and the distribution of the labels and their correlation with the features is initially analysed, plotted or simply counted.
- After the features have been analysed, it is possible to reflect on the validity of the previous null processing and to further optimise the previous data cleaning.

Next Meeting Schedule ---- 5min

Once a week, next week at same room at 2:15pm.

Action Items

Meeting cloesd at 3 PM