**AI-Opening Case Study**

**Artificial Intelligence Supporting Economic Re-Opening Efforts**

**A Case Study Using Predictive Analytics for Risk Profiling of Covid-19 Patients**

Key Steps:

All data files for the case study AI-Opening are posted on LATTE. They are:

- Covid19\_SourceFile.XLSX Patient data file which reports the demographics and diagnosis codes (up to maximum 20 of them)

- 1000AnalyticFileSample.XLSX Sample analytic file with first 1000 patients. You need to build the complete file of your own based on the patient file.

- ICD10\_Dx\_TaxonomyTree.XLSX Excel file with the ICD-10 classification tree we discussed in class. The file also has Gender and Age codes in separate tabs. We call this a Diagnosis Taxonomy. You map every ICD10 Dx code into DGL\_3\_Extend clinical chapter. The ICD-10 codes are over 69,000 while the DGL3 clinical chapters are just over 200 of them.

In order to get from the Source file to Analytic file you will do:

1. Mapping your Age and Sex variables according to the instructions in the Excel file ICD10\_Dx\_TaxonomyTree.XLSX
2. Creating binary dummy variables for your Age and Sex classes.
3. Mapping your DX1 to DX20 health condition variables to DGL\_3\_Extend using the taxonomy tree given in the ICD10\_Dx\_TaxonomyTree.XLSX
4. Create the 229 variables named as entries in the DGL\_3\_Extend. For example the first health condition binary variable will be named BEHAV\_AdjustmentStress and the last one (229th binary variable) will be named UROLG\_GEN\_Status2.
5. In your analytic file any of the 229 binary variables will be 1 if the pt had at least one health condition listed in any of the 20 positions of DX1 to Dx20 that belongs to that clinical chapter. For example starting with the first binary health condition variable BEHAV\_AdjustmentStress in your analytic file, if a pt had at least one of the 14 codes listed below regardless of their appearance in any of the Dx1 to Dx20 positions in the source file you turn the value of her BEHAV\_AdjustmentStress binary variable to 1. If the pt had none of the 14 Behavioral Adjustment Stress related codes then the binary variable will be zero. You develop your code to populate all the 229 variables with 0 (no health condition reported) or 1 (at least one health condition within the clinical chapter reported).

|  |  |
| --- | --- |
| **ICD10 Dx Code** | **ICD10 Dx Code Desc** |
| F4320 | Adjustment disorder, unspecified |
| F4322 | Adjustment disorder with anxiety |
| F4323 | Adjustment disorder with mixed anxiety and depressed mood |
| F4324 | Adjustment disorder with disturbance of conduct |
| F4325 | Adjustment disorder w mixed disturb of emotions and conduct |
| F4329 | Adjustment disorder with other symptoms |
| F438 | Other reactions to severe stress |
| F439 | Reaction to severe stress, unspecified |
| F430 | Acute stress reaction |
| F4310 | Post-traumatic stress disorder, unspecified |
| F4311 | Post-traumatic stress disorder, acute |
| F4312 | Post-traumatic stress disorder, chronic |
| F930 | Separation anxiety disorder of childhood |
| R457 | State of emotional shock and stress, unspecified |

1. When your analytic file is created you can decide to run it with regression models or AI/ML models.
2. With the Mortality variable as the dependent variable of your model (0= pt survived Covid-19 and 1= patient died) follow the slides shared in LATTE and discussed in class to proceed with your case study and come up with the most important risk factors as predictors of the death.

A more detailed version of this mini instruction will be available on LATTE soon. However this mini version and the files and the recording of the class should provide you with enough guidance to get started.