**Document of MS project**

1. **using old edss score cut-off categories variables + Composite cohort**
2. prepare cohort data for descriptive stats and create descriptive stats
3. results -- F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\03 Results
4. (the cohort data for descriptive stats): 2016-07-11 20.01.50
5. (descriptive stats)： DS\_20160713

Iii. (descriptive stats tables for delivery): Lichao added

1. code -- F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\02 Code
2. (prepare the cohort data for descriptive stats): 01 prepare cohort for descriptive stats
3. Jie’s version: Change the input directory (i.e. line 12 of the code “F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\02 Code\01 prepare cohort for descriptive stats\createJie'sVersion\Scripts\ main\_createCohort.R”) into: " F:\\Jie\\MS\\04\_Delivery\\01\_DescriptiveStats\\01 Data\\"

This input is the raw data from Ray

1. Lichao’s verion: Change the input directory (i.e. line 7 of the code “F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\02 Code\01 prepare cohort for descriptive stats\createLichao'sVersion\ main\_lichao\_Jul06.R”) into:

"F:\\Jie\\MS\\04\_Delivery\\01\_DescriptiveStats\\03 Results\\ 2016-07-08 02.16.04\\"

this input is the result of Jie’s version (refer to the code “F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\02 Code\01 prepare cohort for descriptive stats\createJie'sVersion\Scripts\ main\_createCohort.R”)

1. (create the descriptive stats): 02 create descriptive stats

Copy the whole folder (i.e. 2016-07-11 20.01.50 - the result of the Lichao’s version which is based on the code , “F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\02 Code\01 prepare cohort for descriptive stats\createLichao'sVersion\ main\_lichao\_Jul06.R”) to the following folder in the master node.

“s3://emr-rwes-pa-spark-dev-datastore/Jie/MS/Jul12/”

1. Preparing model data for Cmp cohorts
2. results -- F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\03 Results
3. code -- F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\02 Code
4. generate the data for model

change the directory of input file (i.e. line 99 of the code “F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\02 Code\main.R”) into:

“F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\03 Results\2016-07-11 20.01.50\”

This input is the result of the cohort data of Lichao’s version for the descriptive stats

Change the directory of the input file (i.e. line 118 of the code “F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\02 Code\main.R”) into:

“F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\03 Results\2016-07-08 02.16.04”

The input is the result of cohort data of Jie’s version for descriptive stats

1. QC the data for model in the last step
2. change the input directory (i.e. line 5 of the code “F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\02 Code\QcData4ModelUsingRayData.R”) into :

“F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\03 Results”

This input is the result of the cohort data for model

1. change the input directory (i.e. line 7 of the code “F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\02 Code\QcData4ModelUsingRayData.R”) into:

“F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\01 Data”

This input is the raw data from Ray

1. InitModel and nonRegularizedGLM
2. results – F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\03 Results
3. (InitModel with all the base variables): 2016-07-14 20.48.49
4. (InitModel with top 10 variables): 2016-07-15 02.33.53
5. (GLM with top 10 variables): 2016-07-14 22.26.06
6. code -- F:\Jie\MS\04\_Delivery\13\_InitModel\_NonRegulatizedGLM\02 Code
7. initModel with the based variables:

change the input directory (i.e. line 18 of the code “F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\02 Code\InitModel\scripts\ run\_\_BooleanPredictor.R”) into :

“F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\01 Data\2016-07-12 15.43.48”

The input data is the result of “Preparing model data for Cmp cohorts”

Change the line 32 into “main.arglist$bTopVarsOnly <- F”

1. initModel with the top 10 variables selected from the above model.

change the input directory (i.e. line 18 of the code “F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\02 Code\InitModel\scripts\ run\_\_BooleanPredictor.R”) into:

“F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\01 Data\2016-07-12 15.43.48”

The input data is the result of “Preparing model data for Cmp cohorts”

Change the line 32 into “main.arglist$bTopVarsOnly <- F”

Change the input directory (i.e. line 35 of the code “F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\02 Code\InitModel\scripts\ run\_\_BooleanPredictor.R”) into:

“F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\03 Results\2016-07-14 20.48.49”

This input is the results of initModel with based variables

1. Non-regularized glm with the top 10 variables.

Change the input directory (i.e. line 36 of the code “F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\02 Code\ Non\_RegularizedGLM\scripts\main.R”) into :

F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\03 Results\2016-07-15 02.33.53

This input is the results of the InitModel with the to 10 variables

1. Results in Step 3) above put into tables for the composite cohort
2. results –
3. F:\Jie\MS\04\_Delivery\04\_FinalTables\03 Results\2016-07-14 05.33.59
4. tables for delivery: Lichao added
5. code -- F:\Jie\MS\04\_Delivery\04\_FinalTables\02 Code

change the input directory (i.e. line 4 of the code “F:\Jie\MS\04\_Delivery\04\_FinalTables\02 Code\createFinalTables.R) into:

“F:\Jie\MS\04\_Delivery\04\_FinalTables\01 Data\2016-07-14 20.48.49”

This input is the results of InitModel with base variables

change the input directory (i.e. line 7 of the code “F:\Jie\MS\04\_Delivery\04\_FinalTables\02 Code\createFinalTables.R” ) into:

“F:\Jie\MS\04\_Delivery\04\_FinalTables\01 Data\2016-07-15 02.33.53”

This input is the results of InitModel with top 10 variables

change the input directory(i.e. line 9 of the code “F:\Jie\MS\04\_Delivery\04\_FinalTables\02 Code\createFinalTables.R”) into:

“F:\Jie\MS\04\_Delivery\04\_FinalTables\01 Data\ 2016-07-14 22.26.06”

This input is the results of non-regularized glm with top 10 variables

Change the input directory (i.e. line 11 of the code “F:\Jie\MS\04\_Delivery\04\_FinalTables\02 Code\createFinalTables.R”) into:

“F:\Jie\MS\04\_Delivery\04\_FinalTables\01 Data”

This input is the directory of the lookup table of variables this version used.

Change the input file name (i.e. line 12 of the code “F:\Jie\MS\04\_Delivery\04\_FinalTables\02 Code\createFinalTables.R”) into：

“lookup\_20160714.csv”

This input is the file name of the lookup table of the variables for this version.

1. **using new edss score cut-off categories variables + Composite cohort**
2. prepare cohort data for descriptive stats and create descriptive stats
3. results -- F:\Jie\MS\04\_Delivery\11\_DescriptiveStats\03 Results
4. (the cohort data for descriptive stats): 2016-07-19 01.52.43
5. (descriptive stats)： DS\_20160719
6. (descriptive stats tables for delivery): Lichao added
7. code -- F:\Jie\MS\02\_Code\Code4Delivery\11\_DescriptiveStats
8. (prepare the cohort data for descriptive stats): 01 prepare cohort for descriptive stats
9. Jie’s version: Change the input directory (i.e. line 12 of the code “F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\02 Code\01 prepare cohort for descriptive stats\createJie'sVersion\Scripts\main\_createCohort.R”) into: " F:\\Jie\\MS\\04\_Delivery\\01\_DescriptiveStats\\01 Data\\"

The input is the raw data from Ray

1. Lichao’s verion: Change the input directory (i.e. line 6 of the code “F:\Jie\MS\04\_Delivery\11\_DescriptiveStats\02 Code\01 prepare cohort for descriptive stats\createLichao'sVersion\main.R”) into:

"F:\\Jie\\MS\\04\_Delivery\\11\_DescriptiveStats\\03 Results\\ 2016-07-08 02.16.04\\"

1. the input is the result of Jie’s version based on the code in the last step (create the descriptive stats): 02 create descriptive stats

Copy the whole folder (i.e. 2016-07-19 01.52.43) to the following folder in the master node.

“s3://emr-rwes-pa-spark-dev-datastore/Jie/MS/Jul19/2016-07-19 01.52.43/”

The input is the result of Licha’s version based on the code in the last step

1. Preparing model data for Cmp cohorts
2. results -- F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\03 Results
3. The model data after removing references: 2016-07-20 06.58.41
4. The model data after merging some categories: 2016-07-26 04.08.00
5. code -- F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\02 Code
6. generate the data for model

change the directory and file name of input file (i.e. line 6 of the code “F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\02 Code\scripts\GenDataFromRaw.R”) into:

“F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\01 Data\ MS\_decsupp\_analset\_20160701.csv”

This input is the raw data from Ray

Change the directory and file name of the input file (i.e. line 9 of the code “F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\02 Code\ scripts\GenDataFromRaw.R”) into:

“F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\03 Results\cmp4Model.csv”

This input is the result of the cohort data for model, please note that the only useful information for this input is the record\_num which is used to select cohort.

1. merge some categories

change the directory of the input file (i.e. line 4 of the code “F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\02 Code\

mergeCategory.R” ) into:

“F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\03 Results\2016-07-20 06.58.41”

This input is the cohort data from the last step

1. QC the data for model in the last step
2. change the directory of input file (i.e. line 99 of the code “F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\02 Code\main.R”) into:

“F:\Jie\MS\04\_Delivery\11\_DescriptiveStats\03 Results\2016-07-19 01.52.43\”

This input is the result of the cohort data of Lichao’s version for the descriptive stats

1. Change the directory of the input file (i.e. line 118 of the code “F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\02 Code\main.R”) into:

“F:\Jie\MS\04\_Delivery\11\_DescriptiveStats\03 Results\2016-07-08 02.16.04”

The input is the result of cohort data of Jie’s version for descriptive stats

1. change the input file (i.e. line 223 of the code “F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\02 Code\main.R”) into:

“F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\01 Data\ MS\_decsupp\_analset\_20160701.csv”

This input is the raw data from Ray

1. change the directory of the input file (i.e. line 4 of the code “F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\02 Code\

mergeCategory.R” ) into:

“F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\03 Results\2016-07-20 06.58.41”

This input is the cohort data from the last step

1. Run “QC4ModelDataBasedOnGenDataFromRaw.R”(in “F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\02 Code”):

Change the input directory (i.e. line 4) into: “F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\03 Results\2016-07-20 06.58.41“

This input is the model data using the latest method(based on genDataFromRaw.R) and without category merge.

1. Change the input directory (i.e. line 5) into:

F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\03 Results\2016-07-20 08.27.48

This input is the model data using the previous method(based on main.R) and without the category merge.

If the qc code returns “TRUE”, then it means that the qc is passed.

1. InitModel and nonRegularizedGLM
2. Results –F:\Jie\MS\04\_Delivery\13\_InitModel\_NonRegulatizedGLM\03 Results
3. (InitModel with all the base variables): 2016-07-26 04.15.57
4. (InitModel with top 10 variables): 2016-07-26 08.17.58
5. (GLM with top 10 variables): 2016-07-27 04.16.35
6. code -- F:\Jie\MS\04\_Delivery\13\_InitModel\_NonRegulatizedGLM\02 Code
7. initModel with the based variables:

change the input directory (i.e. line 18 of the code “F:\Jie\MS\04\_Delivery\13\_InitModel\_NonRegulatizedGLM\02 Code\InitModel\scripts\ run\_\_BooleanPredictor.R”) into :

“F:\Jie\MS\04\_Delivery\13\_InitModel\_NonRegulatizedGLM\01 Data\2016-07-26 04.08.00”

This input is the model data from the last step (prepare cohort data for model)

Change the line 32 into “main.arglist$bTopVarsOnly <- F”

1. initModel with the top 10 variables selected from the above model.

change the input directory (i.e. line 18 of the code “F:\Jie\MS\04\_Delivery\13\_InitModel\_NonRegulatizedGLM\02 Code\InitModel\scripts\ run\_\_BooleanPredictor.R”) into :

“F:\Jie\MS\04\_Delivery\13\_InitModel\_NonRegulatizedGLM\01 Data\2016-07-26 04.08.00”

Change the line 32 into “main.arglist$bTopVarsOnly <- T”

Change the input directory (i.e. line 36 of the code “F:\Jie\MS\04\_Delivery\13\_InitModel\_NonRegulatizedGLM\02 Code\InitModel\scripts\ run\_\_BooleanPredictor.R”) into:

“F:\Jie\MS\04\_Delivery\13\_InitModel\_NonRegulatizedGLM\03 Results\2016-07-26 04.15.57”

This input is the result of InitModel with base variables based on the code in the last step.

1. Non-regularized glm with the top 10 variables.

Change the input directory (i.e. line 36 of the code “F:\Jie\MS\04\_Delivery\13\_InitModel\_NonRegulatizedGLM\02 Code\ Non\_RegularizedGLM\scripts\main.R”) into :

F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\03 Results\2016-07-26 08.17.58

This input is the result of InitModel with top 10 variables based on the code in the last step

1. Results in Step 3) above put into tables for the composite cohort

|  |
| --- |
| 1. results – 2. F:\Jie\MS\04\_Delivery\14\_FinalTables\03 Results 3. tables for delivery: Lichao added |
| 1. code -- F:\Jie\MS\04\_Delivery\14\_FinalTables\02 Code |

change the input directory (i.e. line 4 of the code “F:\Jie\MS\04\_Delivery\14\_FinalTables\02 Code\createFinalTables.R) into:

“F:\Jie\MS\04\_Delivery\14\_FinalTables\01 Data\2016-07-26 04.15.57”

This input is the result of InitModel with base variables based on the code in

change the input directory (i.e. line 7 of the code “F:\Jie\MS\04\_Delivery\14\_FinalTables\02 Code\createFinalTables.R” ) into:

“F:\Jie\MS\04\_Delivery\14\_FinalTables\01 Data\2016-07-26 08.17.58”

This input is the result of the InitModel with the top 10 variables based on the code in

change the input directory(i.e. line 9 of the code “F:\Jie\MS\04\_Delivery\04\_FinalTables\02 Code\createFinalTables.R”) into:

“F:\Jie\MS\04\_Delivery\14\_FinalTables\01 Data\2016-07-27 04.16.35”

This input is the result of non-regularized glm with the top 10 variables based on the code in

Change the input directory (i.e. line 11 of the code “F:\Jie\MS\04\_Delivery\14\_FinalTables\02 Code\createFinalTables.R”) into:

“F:\Jie\MS\04\_Delivery\14\_FinalTables\01 Data”

This input is the directory of the lookup table of variables this version used.

Change the input file name (i.e. line 12 of the code “F:\Jie\MS\04\_Delivery\14\_FinalTables\02 Code\createFinalTables.R”) into：

“lookup\_20160720.csv”

This input is the file name of the lookup table of the variables for this version.

1. **using new edss score cut-off categories variables + 4 other cohorts**
2. Preparing model data for the other 4 cohorts
3. results -- F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\03 Results
4. code -- F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\02 Code
5. generate the data for model

change the directory and file name of input file (i.e. line 5 of the code “F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\02 Code\scripts\GenDataFromRaw.R”) into:

“F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\01 Data\ MS\_decsupp\_analset\_20160701.csv”

This input is the raw data from Ray

Change the directory of the input file (i.e. line 35 of the code “F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\02 Code\ scripts\GenDataFromRaw.R”) into:

“F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\01 Data\2016-07-25 09.29.06”

This input is the model data based on the previous method –“F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\02 Code\main.R”. Please note that the only useful information of this input is the record\_num which is used to select cohort

1. QC
2. change the directory of input file (i.e. line 120 of the code “F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\02 Code\main.R”) into:

“F:\Jie\MS\04\_Delivery\11\_DescriptiveStats\03 Results\2016-07-19 01.52.43\”

This input is the result of the cohort data of Lichao’s version for the descriptive stats

1. change the input file (i.e. line 128 of the code “F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\02 Code\main.R”) into:

“F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\01 Data\ MS\_decsupp\_analset\_20160701.csv”

This input is the raw data from Ray

1. Change the directory of the input file (i.e. line 140 of the code “F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\02 Code\main.R”) into:

“F:\Jie\MS\04\_Delivery\11\_DescriptiveStats\03 Results\2016-07-08 02.16.04”

The input is the result of cohort data of Jie’s version for descriptive stats

1. Run “QC4ModelDataBasedOnGenDataFromRaw.R”(in “F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\02 Code”):

Change the input directory (i.e. line 4) into: “F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\03 Results\ 2016-08-08 05.39.00 “

This input is the model data using the latest method (based on genDataFromRaw.R) and without category merge.

1. Change the input directory (i.e. line 5) into:

F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\03 Results\ 2016-08-08 06.41.05

This input is the model data using the previous method(based on main.R) and without the category merge.

There is a for loop (each for one cohort) when check if the qc passed.

If the qc code returns 5 “TRUE”, then it means that the qc is passed.

1. InitModel and nonRegularizedGLM
2. Results –F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\03 Results
3. (InitModel with all the base variables): 2016-08-08 08.19.05
4. (InitModel with top 10 variables): 2016-08-08 09.24.44
5. (GLM with top 10 variables): 2016-08-11 06.51.00
6. (InitModel with all the base variables + B2B): 2016-08-11 12.00.36
7. (InitModel with top 10 variables + B2B): 2016-08-11 12.09.42
8. (GLM with top 10 variables + B2B): 2016-08-11 12.14.54
9. code -- F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\02 Code
10. initModel with the based variables: (for B2Fir, B2Sec and BConti)

use line 25 instead of line 26

change the input directory (i.e. line 18 of the code “F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\02 Code\InitModel\scripts\ run\_\_BooleanPredictor.R”) into :

“F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\01 Data\2016-08-08 05.39.00”

This input is the model data which is based on the code in ….

Change the line 36 into “main.arglist$bTopVarsOnly <- F”

1. initModel with the top 10 variables selected from the above model. (for B2Fir, B2Sec and BConti)

use line 25 instead of line 26

change the input directory (i.e. line 7 of the code “F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\02 Code\InitModel\scripts\ run\_\_BooleanPredictor.R”) into:

“F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\01 Data\2016-08-08 05.39.00”

This input is the model data which is based on the code in

Change the line 36 into “main.arglist$bTopVarsOnly <- T”

Change the input directory (i.e. line 39 of the code “F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\02 Code\InitModel\scripts\ run\_\_BooleanPredictor.R”) into:

“F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\03 Results\2016-08-08 08.19.05”

This input is the result of InitModel with the base variables based on the code in

1. Non-regularized glm with the top 10 variables. (for B2Fir, B2Sec and BConti)

use line 25 instead of line 26

Change the input directory (i.e. line 36 of the code “F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\02 Code\ Non\_RegularizedGLM\scripts\main.R”) into :

F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\03 Results\2016-08-08 09.24.44

This input is the result of InitModel with top 10 variables based on the code in

1. initModel with the based variables (especially for B2B):

use line 26 instead of line 25

change the input directory (i.e. line 18 of the code “F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\02 Code\InitModel\scripts\ run\_\_BooleanPredictor.R”) into :

“F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\01 Data\2016-08-08 05.39.00”

This input is the model data based on the code in

Change the line 36 into “main.arglist$bTopVarsOnly <- F”

1. initModel with the top 10 variables selected from the above model. (especially for B2B)

Use line 26 instead of 25

change the input directory (i.e. line 7 of the code “F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\02 Code\InitModel\scripts\ run\_\_BooleanPredictor.R”) into:

“F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\01 Data\2016-08-08 05.39.00”

Change the line 36 into “main.arglist$bTopVarsOnly <- F”

Change the input directory (i.e. line 39 of the code “F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\02 Code\InitModel\scripts\ run\_\_BooleanPredictor.R”) into:

“F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\03 Results\ 2016-08-11 12.00.36”

1. Non-regularized glm with the top 10 variables. (especially for B2B)

Change the input directory (i.e. line 36 of the code “F:\Jie\MS\04\_Delivery\0

3\_InitModel\_NonRegulatizedGLM\02 Code\ Non\_RegularizedGLM\scripts\main.R”) into :

F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\03 Results\ 2016-08-11 12.09.42

1. Results in Step 3) above put into tables for the other 4 cohorts

|  |
| --- |
| 1. results – 2. F:\Jie\MS\04\_Delivery\24\_FinalTables\03 Results\2016-08-10 13.10.01 (corresponding to i, ii, iii above) 3. F:\Jie\MS\04\_Delivery\24\_FinalTables\03 Results\2016-08-11 12.18.09 (corresponding to iv, v, vi above) 4. Tables for delivery: Lichao added |
| 1. code -- F:\Jie\MS\04\_Delivery\24\_FinalTables\02 Code 2. for B2Fir, B2Sec and BConti 3. change the input directory (i.e. line 4 of the code “F:\Jie\MS\04\_Delivery\24\_FinalTables\02 Code\createFinalTables.R) into:   “F:\Jie\MS\04\_Delivery\24\_FinalTables\01 Data\ 2016-08-08 08.19.05”  This input is the result of InitModel with base variables based on the code in   1. change the input directory (i.e. line 7 of the code “F:\Jie\MS\04\_Delivery\24\_FinalTables\02 Code\createFinalTables.R” ) into:   “F:\Jie\MS\04\_Delivery\24\_FinalTables\01 Data\ 2016-08-08 09.24.44”  This input is the result of the InitModel with the top 10 variables based on the code in   1. change the input directory(i.e. line 9 of the code “F:\Jie\MS\04\_Delivery\24\_FinalTables\02 Code\createFinalTables.R”) into:   “F:\Jie\MS\04\_Delivery\24\_FinalTables\01 Data\ 2016-08-11 06.51.00”  This input is the result of non-regularized glm with the top 10 variables based on the code in   1. Change the input directory (i.e. line 13 of the code “F:\Jie\MS\04\_Delivery\24\_FinalTables\02 Code\createFinalTables.R”) into:   “F:\Jie\MS\04\_Delivery\24\_FinalTables\01 Data”  This input is the directory of the lookup table of variables this version used.   1. Change the input file name (i.e. line 14 of the code “F:\Jie\MS\04\_Delivery\24\_FinalTables\02 Code\createFinalTables.R”) into：   “lookup\_20160816.csv”  This input is the file name of the lookup table of the variables for this version.  2. For B2B cohort (removing the variables who can be completely separated by outcome) |
| 1. change the input directory (i.e. line 4 of the code “F:\Jie\MS\04\_Delivery\24\_FinalTables\02 Code\createFinalTables.R) into:   “F:\Jie\MS\04\_Delivery\24\_FinalTables\01 Data\ 2016-08-11 12.00.36”  This input is the result of InitModel with base variables based on the code in   1. change the input directory (i.e. line 7 of the code “F:\Jie\MS\04\_Delivery\24\_FinalTables\02 Code\createFinalTables.R” ) into:   “F:\Jie\MS\04\_Delivery\24\_FinalTables\01 Data\ 2016-08-11 12.09.42”  This input is the result of the InitModel with the top 10 variables based on the code in   1. change the input directory(i.e. line 9 of the code “F:\Jie\MS\04\_Delivery\24\_FinalTables\02 Code\createFinalTables.R”) into:   “F:\Jie\MS\04\_Delivery\24\_FinalTables\01 Data\ 2016-08-11 12.14.54”  This input is the result of non-regularized glm with the top 10 variables based on the code in   1. Change the input directory (i.e. line 11 of the code “F:\Jie\MS\04\_Delivery\24\_FinalTables\02 Code\createFinalTables.R”) into:   “F:\Jie\MS\04\_Delivery\24\_FinalTables\01 Data”  This input is the directory of the lookup table of variables this version used.   1. Change the input file name (i.e. line 12 of the code “F:\Jie\MS\04\_Delivery\24\_FinalTables\02 Code\createFinalTables.R”) into：   “lookup\_20160816.csv”  This input is the file name of the lookup table of the variables for this version. |

Missing:

1. Raw data and dictionary
2. The variable description lookup table.