# Documentation of the Novartis Decision Support Tool Project

**Abstract**: This document records the steps taken to produce deliveries for the Novartis Decision Support Tool project. The objective is to make it easy to reproduce all the deliveries. The document includes the description of data, code, usage and results produced.

1. Descriptive Statistics
   1. Abstract: Two versions of descriptive statistics are delivered: the first with the IMS-proposed cut-offs for EDSS score variables while the second with the Novartis-proposed cut-offs.
   2. With IMS-proposed cut-offs
      1. Preprocess data for computing descriptive statistics
         1. Input: Raw data (data in MS\_decsupp\_analset\_20160701.csv, dictionary in F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\01 Data\Analysis Dataset Variable Description 2016-06-30.xlsx)
         2. The code
            1. General flow: run the ‘automatic’ version (see below) first using the raw data; then run ‘manual correction’ using the results from the ‘automatic’ version. The output is ready for computing the descriptive statistics.
            2. The ‘automatic’ version

Code location: F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\02 Code\01 prepare cohort for descriptive stats\automatic

Code description: It performs initial processing of the data such as dichotomising variables and merging small categories. It can be used to produce two types of results: the first with continuous variables intact and the second with continuous variables dichotomised.

Input: raw data. The input path needs to be changed as follows: 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\automatic\Scripts\main\_createCohort.R”) into: “F:/Jie/MS/04\_Delivery/01\_DescriptiveStats/01 Data/”

How to run the code: execute script ‘./Scripts/main\_createCohort.R’ and it will produce the types of results described in 1.2.1.2.2.2.

Example results: F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\03 Results\2016-07-08 02.16.04. Files ‘\*withoutTransf\*’ are those in which continuous variables are intact while files ‘\*withTransf\*’ are those in which continuous variables are dechotomised.

* + - * 1. ‘Manual Correction’

Note: There’s a bug in this code. The bug-fixed version can be found in the corresponding code in Section 1.3 below. This is updated in the delivery of 1.3.

Code location: F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\02 Code\01 prepare cohort for descriptive stats\manual\_correction

Code description: It performs variable-specific processing to the set of manually-selected ‘base variables’ which will be used for modelling.

Input: both types of results from the ‘automatic’ version above. The input path needs to be changed as follows: 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\manual\_correction\main.R”) into: "F:/Jie/MS/04\_Delivery/01\_DescriptiveStats/03 Results/2016-07-08 02.16.04/"

How to run the code: execute script ‘main.R’

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

* + 1. Compute descriptive statistics
       1. Code location: F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\02 Code\02 create descriptive stats
       2. Input: the results from ‘manual correction’ above. The input paths need to be changed (Lines 17 and 19 of the pyspark code).
       3. How to run the code: execute script ‘run\_DS.sh’ (in the script the Spark code is run 5 times with different arguments, each for a cohort. )
       4. Example results: F:/Jie/MS/04\_Delivery/01\_DescriptiveStats/03 Results/DS\_20160713. Within each cohort’s folder, descriptive statistics are saved in file ‘part-00000’ in the format of csv.
    2. Delivery: \\woksfps01\RWES Central team P&BD\Novartis Clinical Decision Support Tool\00. Overall PMO\04. Meeting materials\2016-07-06.
  1. With Novartis-proposed cut-offs
     1. The code and usage are the same as 1.2 except two changes are made to ‘manual correction’ in 1.2.1.2.3. (the updated code is under F:\Jie\MS\04\_Delivery\11\_DescriptiveStats\02 Code\01 prepare cohort for descriptive stats\manual\_correction). The two changes are:
        1. The bug in 1.2.1.2.3.1 is fixed about dichotomising missing values in continuous variables;
        2. The cut-offs are updated using Novartis-proposed values.
     2. Example results: F:\Jie\MS\04\_Delivery\11\_DescriptiveStats\03 Results\
        1. Data for computing descriptive statistics: 2016-07-19 01.52.43
        2. Descriptive statistics: DS\_20160719
     3. Delivery: \\woksfps01\RWES Central team P&BD\Novartis Clinical Decision Support Tool\00. Overall PMO\04. Meeting materials\2016-08-03\Updated Descriptive Statistics with NVS-proposed EDSS Cut-Offs.

1. Modelling
   1. Abstract: This includes the modelling of the 5 cohorts by 6 outcomes. It also includes different models, namely logistic-regression with the elastic-net penalty and extensive base variables, logistic-regression with the elastic-net penalty and the top 10 most important variables and the standard logistic-regression with the top 10 most important variables. The following divides the documentation into two parts: one for the composite cohort and the other for the other 4 cohorts. This is because firstly the result of the composite cohort is delivered separately prior to that of the other 4 cohorts, and secondly the code for the other 4 cohorts is slightly different for the purpose of incorporating multiple cohorts.
   2. The Composite cohort
      1. Prepare the modelling data
         1. Code location: F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\02 Code\main.R
         2. Code description: The code further pre-processes the data generated for descriptive statistics to generate data ready for modelling (e.g., removing reference categories, etc.)
         3. Input: Outputs from both 1.2.1.2.2 and 1.2.1.2.3 (pre-processed data from the ‘automatic’ version and ‘manual correction’ version, respectively). The input paths need to be changed as the following: 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\", and then 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"
         4. How to run the code: execute the main.R above.
         5. Code for QC the above result:
            1. Location: F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\02 Code\QcData4ModelUsingRawData.R
            2. Input: the raw data and an already-generated dataset for modelling. The input paths need to be changed as follows: Change the already-generated data directory (i.e. line 5 of the code "F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\02 Code\QcData4ModelUsingRawData.R") into: "F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\03 Results" and then change the raw data directory (i.e. line 7 of the code "F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\02 Code \QcData4ModelUsingRawData.R") into: "F:\Jie\MS\04\_Delivery\01\_DescriptiveStats\01 Data"
            3. Usage: Execute the file ‘QcData4ModelUsingRawData.R’. Check the resultant value of bEq at Line 359 and unmatchedRows at Line 366. In order to pass the QC, bEq needs to be True AND unmatchedRows needs to be empty.
         6. Additional variable-merging: after initial modelling, some categories of the baseline EDSS score are merged because they have very similar coefficients and confidence intervals. The code used to do this is in F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\02 Code\mergeCategory.R. The input to this code is the data generated using code described from 2.2.1.1 to 2.2.1.4 (corresponding to variable ‘inDir’ in the code).
         7. Example results:
            1. Using IMS-proposed EDSS cut-offs and with additional variable-merging: F:\Jie\MS\04\_Delivery\02\_ExtractModelData4Cmp\03 Results
            2. Using Novartis-proposed EDSS cut-offs and without additional variable merging: F:\Jie\MS\04\_Delivery\12\_ExtractModelData4Cmp\03 Results\2016-07-20 06.58.41
      2. Modelling
         1. Modelling is done by two pieces of code: InitModel and NonRegularizedGLM.
         2. InitModel
            1. Code location: F:\Jie\MS\04\_Delivery\13\_InitModel\_NonRegulatizedGLM\02 Code\InitModel
            2. Code description: The code can be used to produce two types of results: the logistic regression with the elastic-net penalty and the extensive set of variables and the logistic regression with the elastic-net penalty and the top 10 most important variables.
            3. Input: data for model produced in 2.2.1. If it is for logistic regression with the elastic-net penalty and the top 10 most important variables, the result from logistic regression with the elastic-net penalty and the extensive set of variables is also needed. The paths for the input need to be updated to the correct locations: variable ‘main.arglist$data\_dir’ in “./scripts/run\_\_BooleanPredictor.R” should be the location of the data for model produced in 2.2.1, while variable ‘main.arglist$initEnetDir’ should be the location of the result from logistic regression with the elastic-net penalty and the extensive set of variables.
            4. How to run the code:

To run the logistic regression with the elastic-net penalty and the extensive set of variables, execute “./scripts/run\_\_BooleanPredictor.R” with the variable ‘main.arglist$bTopVarsOnly’ set to ‘FALSE’ and use the corresponding input in 2.2.2.2.3.

To run the logistic regression with the elastic-net penalty and the top 10 most important variables, execute “./scripts/run\_\_BooleanPredictor.R” with the variable ‘main.arglist$bTopVarsOnly’ set to ‘TRUE and use the corresponding input in 2.2.2.2.3.

* + - * 1. Example results: F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\03 Results

Logistic regression with the elastic-net penalty and the extensive set of variables: 2016-07-14 20.48.49 (Lichao to add: explain the results). Results from InitModel all take the same format: files ‘AvAUCs\_train.csv’ and‘AvAUCs\_test.csv’ save the average AUC values across evaluation folds for all cohort and outcome combinations; Within each cohort, results are put into different outcomes; Within the folder of each cohort / outcome combination, file ‘av\_coefs\_Cmp.csv’ is the average coefficient’s magnitude information, ‘av\_ranking\_Cmp.csv’ contains the average ranking of the predictors, ‘Cmp\_data\_for\_model.csv’ contains the data used for modelling for this particular cohort / outcome combination, ‘Cmp\_params.csv’ contains the best hyper-parameter combinations for all evaluation folds, ‘Cmp\_probs.csv’ is the predicted probabilities of all evaluation folds, ‘Cmp\_roc.png’ is the plotted ROC curve, ‘coefs\_Cmp.csv’ is the predictor coefficients in combinations of all evaluation folds and all alpha values (relative weighting between the L1 and L2 regularisation terms), and ‘rankings\_Cmp.csv’ is the rankings of the predictor coefficients in combinations of all evaluation folds and all alpha values.

logistic regression with the elastic-net penalty and the top 10 most important variables: 2016-07-15 02.33.5. The contents in the result folder are the same as that described in 2.2.2.2.5.1 (Lichao to add: explain the results)

* + - 1. NonRegularizedGLM
         1. Code location: F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\02 Code\Non\_RegularizedGLM
         2. Code description: This code is used to produce the AUC of the standard logistic regression with the top 10 most important variables and the confidence intervals and p-values of those variables.
         3. Input: the output of 2.2.2.2.4.2 above (logistic-regression with the top 10 most important variables). The corresponding line to change in the script is Line 36 of “./scripts/main.R”
         4. How to run the code: execute “./scripts/main.R”
         5. Example results: F:\Jie\MS\04\_Delivery\03\_InitModel\_NonRegulatizedGLM\03 Results\2016-07-14 22.26.06 (Lichao to add: explain the results). Results from NonRegularizedGLM all take the same format: file ‘ci\_auc\_glm.csv’ contains the confidence interval of standard logistic regression, ‘ci\_auc\_glmnet.csv’ contains the confidence interval of logistic regression with the elastic-net penalty, the two PNG images are the plots of the ROC curves of the two models, ‘coef\_info.csv’ contains the confidence interval and p-value information of the top-10 most important variables, ‘roc\_compare.txt’ contains results of comparing the ROC curves from both models and ‘selectedVarNames.csv’ is the list of predictor names.
    1. Collecting results
       1. Code location: F:\Jie\MS\04\_Delivery\14\_FinalTables\02 Code
       2. Code description: The code collects results from 2.2.2 above into tables.
       3. Input: results from 2.2.2.2 and 2.2.2.3 above (i.e., logistic regression with the elastic-net penalty and the extensive set of variables, logistic regression with the elastic-net penalty and the top 10 most important variables and standard logistic regression with the top 10 most important variables). The input paths need to be changed (the following is an example):
          1. 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\04\_FinalTables\01 Data\2016-07-14 20.48.49” (logistic regression with the elastic-net penalty and the extensive set of variables),
          2. 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\04\_FinalTables\01 Data\2016-07-15 02.33.53” (logistic regression with the elastic-net penalty and the top 10 most important variables )
          3. change the input directory(i.e. line 9 of the code “F:\Jie\MS\04\_Delivery\14\_FinalTables\02 Code\createFinalTables.R”) into: “F:\Jie\MS\04\_Delivery\04\_FinalTables\01 Data\2016-07-14 22.26.06” (standard logistic regression with the top 10 most important variables)
       4. How to run the code: execute createFinalTables.R.
       5. Example results:
          1. With IMS-proposed EDSS cut-offs: F:\Jie\MS\04\_Delivery\04\_FinalTables\03 Results\2016-07-15 18.46.56. In this folder, Table 1 is the summary of AUCs from all models and outcomes, Table2\* are importance ranking of the extensive variables, Table3\* are the confidence intervals and p-values of the top-10 most important variables, and table 4 is the risk quintiles.
          2. With Novartis-proposed EDSS cut-offs: F:\Jie\MS\04\_Delivery\14\_FinalTables\03 Results\2016-07-27 06.10.28. The content in this folder is similar to 2.2.3.5.1 above.
    2. Delivery: (Lichao to add) (Also note that the tables are not in the same order / indices as those in 2.2.3 above) Note that the tables in the following deliveries are in a different order from that produced in 2.2.3. Each of these delivery tables are self-explained.
       1. With IMS-proposed cut-offs: \\woksfps01\RWES Central team P&BD\Novartis Clinical Decision Support Tool\00. Overall PMO\04. Meeting materials\2016-07-20
       2. With Novartis-proposed cut-offs: \\woksfps01\RWES Central team P&BD\Novartis Clinical Decision Support Tool\00. Overall PMO\04. Meeting materials\2016-08-03
  1. All other cohorts
     1. Code for all other cohorts is similar to that in 2.2.1, 2.2.2 and 2.2.3 for the composite cohort, with small changes to adapt to produce results for multiple cohorts. Note that modelling results for the other cohorts are only from using the Novartis-proposed EDSS cut-offs.
     2. Prepare the modelling data
        1. Code location: F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\02 Code\scripts\GenDataFromRaw.R
        2. Code description: Prepare modelling data for the other 4 cohorts using those for computing descriptive statistics
        3. Input: Raw data and information about which records are to be randomly selected if they are from the same patients in the same cohort (an example dataset containing this information is ). The input paths could be changed in Lines 5 and 35 for the raw dataset and the record-selection information, respectively.
        4. How to run the code: execute the script.
        5. Code for QC:
           1. Code location: Under F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\02 Code\scripts , ./scripts/main.R and QC4ModelDataBasedOnGenDataFromRaw.R are for QC
           2. Code description: ./scripts/main.R generates the data for model using a different method. QC4ModelDataBasedOnGenDataFromRaw.R compares results from ./scripts/main.R and ./scripts/GenDataFromRaw.R mentioned in 2.3.2.1 above.
           3. Input:

./scripts/main.R: the output from the ‘manual correction’ version for descriptive statistics (the corresponding line to change in ./scripts/main.R is Line 120), the raw data (the corresponding line to change in ./scripts/main.R is Line 128) and the output from the automatic version for descriptive statistics (the corresponding line to change in ./scripts/main.R is Line 140).

QC4ModelDataBasedOnGenDataFromRaw.R: the output from ./scripts/GenDataFromRaw.R in 2.3.2.1 above (the corresponding line to change in QC4ModelDataBasedOnGenDataFromRaw.R is Line 4) and the output from ./scripts/main.R in 2.3.2.5.3.1 above (the corresponding line to change in QC4ModelDataBasedOnGenDataFromRaw.R is Line 5)

* + - * 1. How to justify whether the QC is successfully passed: execute QC4ModelDataBasedOnGenDataFromRaw.R, and if it prints 5 ‘TRUE’ on the screen, the QC is passed successfully.
      1. Example results: F:\Jie\MS\04\_Delivery\22\_ExtractModelData4Cmp\03 Results (the two result folders are generated by 2.3.2.1 and 2.3.2.5)
    1. Modelling
       1. InitModel
          1. Code location: F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\02 Code\InitModel (Lichao to do: remove unused files)
          2. Code description: It runs the logistic-regression with the elastic-net penalty. There are 4 different situations: whether to use the extensive set of variables or the top 10 most important variables, and whether it is for the BRACE to BRACE cohort or the other three cohorts (BRACE to Firstline Oral, BRACE to Secondline and BRACE Continuation). The reason that BRACE to BRACE is separated is that there are some complete-separation variables. The following subsections will be described according to these 4 situations.
          3. Input: Similar to what is described in 2.2.2.2.3, if it is for the extensive variables, the input is the model for data obtained from 2.3.2; if it is for the top 10 most important variables, the input is the model for data from 2.3.2 and the corresponding modelling result from using the extensive set of variables. The paths for the input need to be updated to the correct locations: variable ‘main.arglist$data\_dir’ in “./scripts/run\_\_BooleanPredictor.R” should be the location of the data for model produced in 2.2.1, while variable ‘main.arglist$initEnetDir’ should be the location of the result from logistic regression with the elastic-net penalty and the extensive set of variables.
          4. How to run the code: the same as 2.2.2.2.4
          5. Example results: Under F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\03 Results

Extensive variables and BRACE to BRACE: 2016-08-11 12.00.36

Extensive variables and BRACE to Firstline Oral, BRACE to Secondline and BRACE Continuation: 2016-08-08 08.19.05

Top 10 variables and BRACE to BRACE: 2016-08-11 12.09.42

Top 10 variables and BRACE to Firstline Oral, BRACE to Secondline and BRACE Continuation: 2016-08-08 09.24.44

* + - 1. NonRegularizedGLM
         1. Code location: F:\Jie\MS\04\_Delivery\23\_InitModel\_NonRegulatizedGLM\02 Code\Non\_RegularizedGLM\
         2. Code description: see 2.2.2.3.2.
         3. Input: similar to 2.2.2.3.3, the output of logistic regression with the elastic-net penalty and the top 10 most important variables described in 2.3.3.1 above. Line 53 needs to be changed to the correct input path.
         4. How to run the code: similar to 2.2.2.3.4.
         5. Example results:

BRACE to BRACE: 2016-08-11 12.14.54

BRACE to Firstline Oral, BRACE to Secondline and BRACE Continuation: 2016-08-11 06.51.00.

* + 1. Collecting results: Everything is similar to 2.2.3 but with slightly changed code (e.g., changing the table names, etc.). Example results are in F:\Jie\MS\04\_Delivery\24\_FinalTables\03 Results. Note that the table orders are different from those in 2.2.3.5: Table1 is still the summary of AUCs, Table2 is the risk quintile information, Table3\* are the importance ranking of the extensive set of variables and Table4\* are the confidence intervals and p-values of the top-10 most important variables.
    2. Delivery: \\woksfps01\RWES Central team P&BD\Novartis Clinical Decision Support Tool\00. Overall PMO\04. Meeting materials\2016-08-17. In this delivery, tables for each cohort (similar to 2.2.4) are gathered into one excel file. Also note that if a cohort / outcome combination has less than 100 patients with the positive outcome, its modelling result is not included.

1. Exploring New Variables (Not Delivered) to add
   1. Description: This part is for quickly exploring whether there are extra variables with predictive information but not included as predictors in Section 2. The exploration is focused on the composite cohort and three outcomes: relapse, EDSS progression and confirmed EDSS progression. Only half of the data is used for the entire exploration process. The steps for the exploration is as follows: firstly, all the extra variables mentioned in the protocol are put into mutually exclusive groups; secondly, random forests is used to identify for each outcome, how informative every extra variable group is; thirdly, logistic regression with the elastic-net penalty are trained using the top 1 or 2 most informative groups for each outcome are combined with the base variables used in Section 2 to justify the usefulness of the extra variables. For both random forests and logistic-regression, the same half data is used to perform cross-evaluation. The following is the details.
   2. Prepare the modelling data for RF modelling
      1. The model data used here is just the similar as 2.2.7.1.2 but Lichao to complete
      2. Location: F:\Jie\MS\04\_Delivery\32\_RF\01 Data\2016-07-26 04.08.00
   3. Modelling
      1. RF （stratify the model data into 2 parts, and we just use one part to do the modelling, and never touch the other part.）
         1. Code location: F:\Jie\MS\04\_Delivery\32\_RF\02 Code
         2. Code description: It performs extra-variable grouping and runs the Random Forest multiple times. Each time, the entire set of base variables used in Section 2 are used together with one group of extra variables.
         3. Input: there are two input data.
            1. One is the model data from what is described in 3.2.2,
            2. The other is the raw data. Its location is described in 1.2.1.1.
         4. How to run the code:
            1. Change the following accordingly and execute script ‘./scripts/’ to complete
            2. Hyper parameter setting for grid search, including ntree, mtry, sampsize. Please change them in the line 16, 19, 20
            3. Directory of Model data setting, please change it in the line 28
            4. Outcomes vector you want to focus on, please change it in line 32
            5. Raw data input setting, please change it in line 45
            6. Cores usage for each sub-parallelization. Please change it in line 127
            7. Whether using test mode, please note that test mode should only be used for debug. please change it in line 122
         5. Example results: Under F:\Jie\MS\04\_Delivery\32\_RF\03 Results\2016-08-11 12.00.36. It is the summary table of AUCs on training and test data using different groups.
      2. Elastic-net
         1. Prepare model data:
            1. Code location: F:\Jie\MS\04\_Delivery\33\_InitModel\02 Code\01 prepare model data for InitModel
            2. Code description: adding to the data for modelling several new group variables selected based on the modelling results of Random Forest described in 3.3.1.
            3. Input: there are two inputs.

Modelling data with only base variables with Novartise-propsed EDSS cut-offs. Location: F:\Jie\MS\04\_Delivery\33\_InitModel\01 Data\02 InitModel\2016-07-26 04.08.00

Raw data.

* + - * 1. How to run the code:

Change Line 7 for the correct path of modelling data with only base variables (e.g., 3.3.2.1.3.1)

Change Line 13 for the correct path of raw data.

* + - * 1. An example result: F:\Jie\MS\04\_Delivery\33\_InitModel\03 Results\01 prepare model data for InitModel\2016-08-03 09.40.35
      1. Modelling
         1. Code location: F:\Jie\MS\04\_Delivery\33\_InitModel\02 Code\02 InitModel
         2. Code description: Generate performance evaluation using logistic-regression with the elastic-net penalty and the selected variable groups.
         3. Input: there are two iputs

Model data with only base variables with new edss score cut-off points. Location: F:\Jie\MS\04\_Delivery\33\_InitModel\01 Data\02 InitModel\2016-07-26 04.08.00

Model data after adding the new group variables. Location: F:\Jie\MS\04\_Delivery\33\_InitModel\01 Data\02 InitModel\2016-08-03 09.40.35

* + - * 1. How to run the code:

Model data with new group variables added, please change the line 18 (in the code scripts\run\_\_BooleanPredictor.R) into the directory where the data in 3.3.2.2.3.2 saved in.

Change the line 36 (in the code scripts\run\_\_BooleanPredictor.R) into “FALSE” to build model without top 10 variables selection.

* 1. An example result: Under F:\Jie\MS\04\_Delivery\33\_InitModel\03 Results\02 InitModel\2016-08-04 04.41.57

1. The latest variable lookup table
   1. This is a table translating variables used in descriptive statistics and modelling into proper description. (Lichao to add where it is)