

# Package ‘PheValuator’

December 10, 2018

**Type** Package

**Title** Rapid Phenotype Evaluation

**Version** 0.1.0

**Author** Joel N. Swerdel

**Maintainer** Joel N. Swerdel <jswerdel@its.jnj.com>

**Description** A package for evaluating phenotype algorithms. The package develops a diagnostic prediction model using a set of noisy positives and noisy negatives. It can then apply the model to a large cohort to create a set of subjects each with a predicted probability for the health outcome of interest. This cohort can then be used to evaluate phenotype algorithms through the calculation of the necessary performance characteristics - sensitivity, specificity, and positive and negative predictiver value.

**License** Apache License 2.0

**Encoding** UTF-8

**LazyData** true

**Depends** R (>= 3.2.2),  
DatabaseConnector (>= 2.1.4)

**Imports** PatientLevelPrediction (>= 2.0.4),  
FeatureExtraction (>= 2.1.5),  
SqlRender (>= 1.5.2),  
data.table (>= 1.11.8)

**RoxygenNote** 6.1.1

## R topics documented:

createEvalCohort . . . . .	2
createPhenoModel . . . . .	3
testPhenotype . . . . .	4

<b>Index</b>	<b>6</b>
--------------	----------

---

<code>createEvalCohort</code>	<i>createEvalCohort</i>
-------------------------------	-------------------------

---

## Description

Create the evaluation cohort

## Usage

```
createEvalCohort(connectionDetails = list(), xSpecCohort = "",
  cdmDatabaseSchema = "", cohortDatabaseSchema = "",
  cohortDatabaseTable = "", outDatabaseSchema = "", testOutFile = "",
  trainOutFile = "", estPPV = 1, modelAnalysisId = "",
  evalAnalysisId = "1", cdmShortName = "CDM", mainPopnCohort = 0,
  lowerAgeLimit = 0, upperAgeLimit = 120, startDate = "19001010",
  endDate = "21000101")
```

## Arguments

<code>connectionDetails</code>	connectionDetails created using the function <code>createConnectionDetails</code> in the DatabaseConnector package.
<code>xSpecCohort</code>	The number of the "extremely specific (xSpec)" cohort definition id in the cohort table (for noisy positives)
<code>cdmDatabaseSchema</code>	The name of the database schema that contains the OMOP CDM instance. Requires read permissions to this database. On SQL Server, this should specify both the database and the schema, so for example 'cdm_instance.dbo'.
<code>cohortDatabaseSchema</code>	The name of the database schema that is the location where the cohort data used to define the at risk cohort is available. Requires read permissions to this database.
<code>cohortDatabaseTable</code>	The tablename that contains the at risk cohort. The expectation is cohortTable has format of COHORT table: cohort_concept_id, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.
<code>outDatabaseSchema</code>	The name of the database schema that is the location where the data used to define the outcome cohorts is available. Requires read permissions to this database.
<code>testOutFile</code>	A string designation for the evaluation cohort file
<code>trainOutFile</code>	A string designation for the training model file
<code>estPPV</code>	A value between 0 and 1 as an estimate for the positive predictive value for the <code>exclCohort</code>
<code>modelAnalysisId</code>	The string previously used for designating the name for the model files
<code>evalAnalysisId</code>	Another string for designating the name for the evaluation files
<code>cdmShortName</code>	A short name for the current database (CDM)
<code>mainPopnCohort</code>	The number of the cohort to be used as a base population for the model (default=NULL)

lowerAgeLimit	The lower age for subjects in the model (default=NULL)
upperAgeLimit	The upper age for subjects in the model (default=NULL)
startDate	The starting date for including subjects in the model (default=NULL)
endDate	The ending date for including subjects in the model (default=NULL)

## Details

Creates the evaluation cohort and applies a diagnostic prediction model for determination of a probability for the health outcome of interest

---

createPhenoModel	<i>createPhenoModel</i>
------------------	-------------------------

---

## Description

Create the phenotype model

## Usage

```
createPhenoModel(connectionDetails = list(), xSpecCohort = "",
  cdmDatabaseSchema = "", cohortDatabaseSchema = "",
  cohortDatabaseTable = "", outDatabaseSchema = "",
  trainOutFile = "", exclCohort = "", estPPV = 1,
  modelAnalysisId = "1", excludedConcepts = c(),
  cdmShortName = "CDM", mainPopnCohort = 0, lowerAgeLimit = 0,
  upperAgeLimit = 120, startDate = "19000101", endDate = "21000101")
```

## Arguments

connectionDetails	connectionDetails created using the function createConnectionDetails in the DatabaseConnector package.
xSpecCohort	The number of the "extremely specific (xSpec)" cohort definition id in the cohort table (for noisy positives)
cdmDatabaseSchema	The name of the database schema that contains the OMOP CDM instance. Requires read permissions to this database. On SQL Server, this should specify both the database and the schema, so for example 'cdm_instance.dbo'.
cohortDatabaseSchema	The name of the database schema that is the location where the cohort data used to define the at risk cohort is available. Requires read permissions to this database.
cohortDatabaseTable	The tablename that contains the at risk cohort. The expectation is cohortTable has format of COHORT table: cohort_concept_id, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.
outDatabaseSchema	The name of a database schema where the user has write capability. A temporary cohort table will be created here.
trainOutFile	A string designation for the training model file

exclCohort	The number of the "extremely sensitive (xSens)" cohort definition id in the cohort table (used to estimate population prevalence and to exclude subjects from the noisy positives)
estPPV	A value between 0 and 1 as an estimate for the positive predictive value for the exclCohort
modelAnalysisId	Another string for designating the name for the model files
excludedConcepts	A list of conceptIds to exclude from featureExtraction
cdmShortName	A short name for the current database (CDM)
mainPopnCohort	The number of the cohort to be used as a base population for the model (default=NULL)
lowerAgeLimit	The lower age for subjects in the model (default=NULL)
upperAgeLimit	The upper age for subjects in the model (default=NULL)
startDate	The starting date for including subjects in the model (default=NULL)
endDate	The ending date for including subjects in the model (default=NULL)

### Details

Function to create a diagnostic prediction model for a health outcome of interest using the xSpec cohort. The model may be applied to the evaluation cohort to determine probabilities for each subject for the health outcome of interest.

---

testPhenotype	<i>testPhenotype</i>
---------------	----------------------

---

### Description

Test phenotype algorithms

### Usage

```
testPhenotype(connectionDetails = list(), cutPoints = c(0.1, 0.2, 0.3,
  0.4, 0.5, "EV", 0.6, 0.7, 0.8, 0.9), resultsFileName = "",
  cohortPheno = "", phenText = "Test Pheno", order = 1,
  testText = "Phenotype", cohortDatabaseSchema = "",
  cohortTable = "", estPPV = 1, cdmShortName = "CDM")
```

### Arguments

connectionDetails	ConnectionDetails created using the function createConnectionDetails in the DatabaseConnector package.
cutPoints	A list of threshold predictions for the evaluations. Include "EV" for the expected value
resultsFileName	The full file name with path for the evaluation file
cohortPheno	The number of the cohort of the phenotype algorithm to test

phenText	A string to identify the phenotype algorithm in the output file
order	The order of this algorithm for sorting in the output file (used when there are multiple phenotypes to test)
testText	Descriptive name for the model
cohortDatabaseSchema	The name of the database schema that is the location where the cohort data used to define the at risk cohort is available. Requires read permissions to this database.
cohortTable	The tablename that contains the at risk cohort. The expectation is cohortTable has format of COHORT table: cohort_concept_id, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.
estPPV	The positive predictive value estimate used in developing the model
cdmShortName	A short name for the current database (CDM)

### Details

This function will perform the phenotype algorithm evaluation using the evaluation cohort returned from createEvalCohort and the phenotype algorithm cohort specified

### Value

A list containing 2 dataframes: 1) results - a dataframe with the results from the phenotype algorithm evaluation 2) misses - a dataframe with a sample of subject ids for TPs, FPs, TNs, and FNs for the 50 percent and over prediction threshold

# Index

`createEvalCohort`, [2](#)

`createPhenoModel`, [3](#)

`testPhenotype`, [4](#)