

ICU Extraktionstool R-Studio

Installation

Um das ICU Extraktionstool R-Studio zu installieren, müssen zunächst R und RStudio installiert werden. Sie können sie von den folgenden Links herunterladen:

R: <https://cran.uni-muenster.de/bin/windows/base/R-3.5.1-win.exe>

Rtools: <https://cran.uni-muenster.de/bin/windows/Rtools/Rtools35.exe>

RStudio: <https://download1.rstudio.org/RStudio-1.1.456.exe>

RStudio ist eine grafische Benutzeroberfläche für das Statistik Tool.

Projekt einrichten

Nachdem Sie R und RStudio installiert haben, können Sie das Archiv des ICU Extraktionstools R-Studio an einen beliebigen Ort entpacken.

Öffnen Sie RStudio und navigieren Sie zum entpackten Verzeichnis im Bereich "File Browser". Hier finden Sie eine Datei namens report.Rmd, eine Datei namens README und einen Ordner namens scripts.

Klicken Sie auf "More" (das grüne Zahnrad) und wählen Sie "SET as working directory" / "Als Arbeitsverzeichnis setzen" aus, um das Verzeichnis als Arbeitsverzeichnis festzulegen.

Öffnen Sie report.Rmd und passen Sie in Zeile 63/64 die beiden Pfade ddata (wo sich die XLSX-Dateien befinden) und fout (wo die Daten gespeichert werden sollen) an. Speichern Sie die Datei.

Ausführen

Klicken Sie schließlich auf den kleinen Pfeil oben neben "Knit" und wählen Sie "knit to HTML", um das Tool auszuführen.

Funktion des Codes

Der Code in diesem Abschnitt dient der Vorverarbeitung von Patientendaten, die in mehreren Excel-Dateien vorliegen, indem er die erforderlichen Informationen aus ihnen extrahiert und in einer einzigen Tabelle zusammenführt. Bei der Vorverarbeitung werden zwei unterschiedliche Schlüssel erzeugt: key_visit und key_patient, wobei key_visit für jeden Eintrag und key_patient für jeden Patienten eindeutig ist.

Die Vorverarbeitung erfolgt in drei Hauptabschnitten, in denen die Daten aus jedem Blatt der Excel-Dateien extrahiert werden. Die drei Abschnitte sind:

Stammdaten: Dieses Blatt enthält die primären Informationen über jeden Patienten, und die meisten Daten werden unverändert extrahiert. Für die komplexen Daten wie "Diagnose_Behandlung", "Anamnese", "Medikation" usw. werden alle Einträge unter den jeweiligen Überschriften extrahiert und zu einer einzigen Zeichenfolge verkettet. Alle Einträge werden zu einem einzigen Wert für den jeweiligen Patienten zusammengeführt.

Euroscore II: Die Daten in diesem Blatt werden so extrahiert, wie sie sind, und die Namen werden für jeden Eintrag erstellt. Als Werte wird der Eintrag im jeweiligen Feld der Spalte "C" hinzugefügt. Der Mortalitätsrisikoscore wird direkt aus dem Feld D47 extrahiert.

Verlegungskriterien: Ähnlich wie im Blatt Euroscore II werden die Namen für jeden Eintrag erstellt, und als Werte wird ein Tripel der Form "JA;;TEILWEISE;;NEIN:" erstellt, und der Wert im entsprechenden Feld des Blattes wird zu jedem Wert des Tripels geschrieben. Die letzten drei Zeilen (43-45) werden nach demselben Konzept extrahiert, jedoch mit anderen Tripeln (d. h. "empfohlen;;erfolgt;;nicht erfolgt:").

Nach der Vorverarbeitung werden die extrahierten Informationen in einer einzigen Tabelle zusammengeführt, die dann als TSV-Datei gespeichert wird. Der Code ist so aufgebaut, dass er alle Excel-Dateien in einem bestimmten Verzeichnis liest, die Daten vorverarbeitet und die zusammengeführte Tabelle als TSV-Datei speichert.