

前列腺癌資料及診斷預測模型

Prostate cancer diagnosis prediction model

指導教授：蔣榮先

專題成員：張嘉澤、王宥富

開發工具：R version 3.1.3、Python 2.7.6

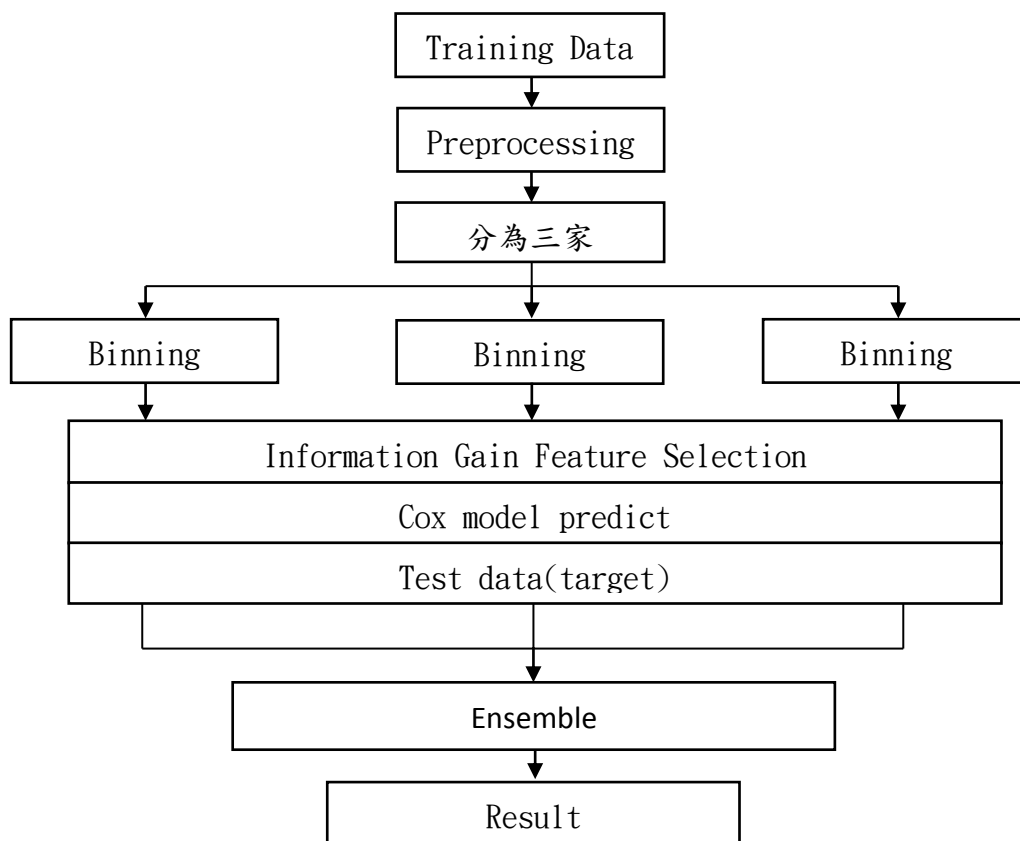
測試環境：Linux、Windows 8

一. 簡介：

這項研究主要是希望針對病人的資料(並不限於前列腺癌)做整理並透過各種演算法或模型給予醫生在病人的死亡率和存活天數上建議，以便協助醫生評估病人的狀況並判斷該如何進行下一步的治療。

取得病人資料後，首先對其作資料的前處理，接著根據不同的醫院分為不同的資料堆(此處是三家)。對各資料堆分別進行離散化(entropy-based binning)、特徵選取(information gain)，以及套用 Cox 生存模型預測，產生出的三個結果統整再對其作整合處理以產生出最終結果。

以下為預測模型架構圖

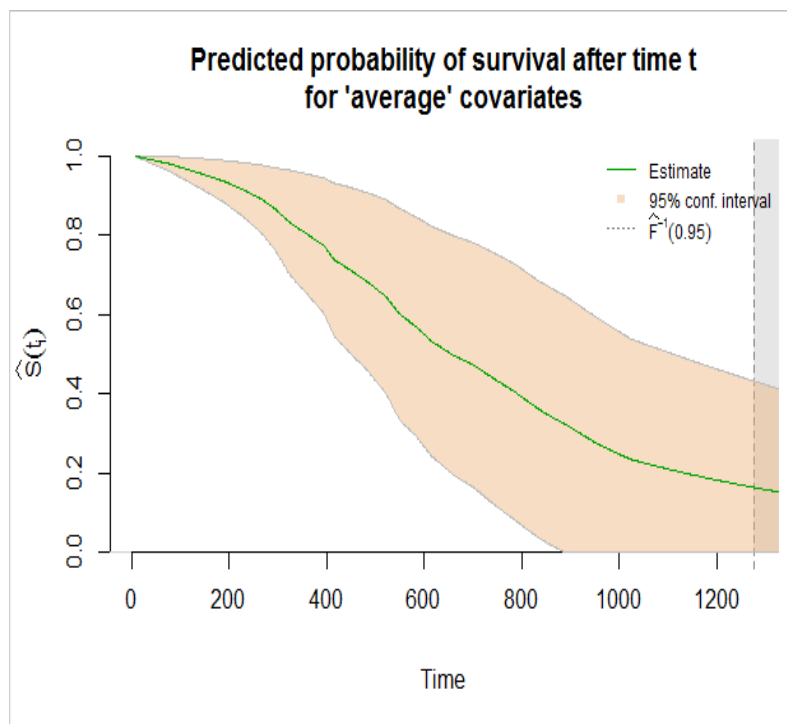


二. 測試結果：

以下為套用此模型對前列腺癌患者分析出來的結果，左圖的 RPT 為病人代號，Risk 代表的則是其死亡風險的高低。右圖為預測患者接下來以時間為變量的平均生存機率曲線，以及 95 百分比信賴區間。

醫生可以藉由參考該資料來初步了解患者的病況，或者根據情形對其用藥做出調整。

RPT	RISK
AZ-00002	32.16667
AZ-00004	146.8333
AZ-00009	146.8333
AZ-00015	91
AZ-00017	78
AZ-00018	105.6667
AZ-00021	141.3333
AZ-00022	99.33333
AZ-00023	70.33333
AZ-00026	132.5
AZ-00028	61.83333
AZ-00032	146.8333
AZ-00039	112.3333
AZ-00041	55.16667
AZ-00042	114.8333
AZ-00045	68.83333
AZ-00046	95.33333
AZ-00047	88.83333
AZ-00051	16.33333



我們希望透過此研究方法，藉由機器學習法判斷病人的狀態，利用大量的資料訓練電腦，期望能達到預估準確的目的性，以幫助醫生能更有效的去做治療。將醫生的專業知識領域結合電腦學習，以提升醫療品質和效率，減少人力資源之浪費。