

Applying data mining algorithm on Al-hakim CBC data to prepare for a disease's prediction application

Group Member:

(Aya Ramadan, Lara Break, Dania Taqtouq, Ikhlas Amer)

Project Abstract

The main objective of the project is to prepare for an automated medical system that predicts diseases, by applying data mining on the complete blood count (CBC) data of Al-Hakim Medical Complex, initially, we have prepared the data and cleaned it, performed some descriptive analyzes, and started applying the data mining algorithms.

Where the association rule was applied to find relationships between tests, and the clustering algorithm was applied, which included several methods for finding the number of clusters, such as the (Elbow method & Silhouette method), and then the last algorithm was applied which is the classification to identify each patient in the data set to which cluster belong, and each cluster has test results that differ from the other, which were determined based on the classification results, and thus preparations were made for the establishment of the medical system.

ملخص المشروع

الهدف الرئيسي من المشروع التحضير لنظام طبي آلي يتنبأ بالأمراض, وذلك من خلال تطبيق تنقيب البيانات على بيانات فحص الدم الشامل (CBC) الخاصة بمجمع الحكيم الطبي, حيث قمنا بتحضير البيانات وتنظيفها وإجراء بعض التحليلات الوصفية, والمباشرة بتطبيق خوارزميات تنقيب البيانات .

حيث تم تطبيق قواعد الإرتباط (Association rule) لإيجاد علاقات بين الفحوصات, وتطبيق خوارزمية التجميع (clustering) التي تضمنت طرقاً عدة لإيجاد عدد المجموعات مثل (Elbow method & Silhouette method), ومن ثم تطبيق آخر خوارزمية وهي التصنيف (Classification) لتحديد كل مريض في مجموعة البيانات لأي كلبتر ينتمي وكل كلبتر له نتائج فحوصات تختلف عن الاخر تم تحديدها بناء على نتائج التصنيف, وبذلك تم التحضير من أجل إنشاء النظام الطبي.