UNIVERSITY OF SCIENCE AND TECHNOLOGY
COLLEGE OF GRADUATE STUDIES AND ACADEMIC ADVANCEMENT

Implementation of Decision Tree Algorithm Using J48 Classifier For Predict Heart Disease

A Thesis

Submitted to the College of Graduate Studies and Academic Advancements

In Partial Fulfillment of the Requirement for the Degree of Master of Science in Information Technology

Prepared By:

Rabah Mohamed Osman Khatmi Mahmoud

Supervisor By:

Dr. Atif Ali Mohamed

January 2017

Abstract

Data mining is the process of discovering interesting patterns and knowledge from large amounts of data. The data sources can include databases, data warehouses, the Web, other information repositories, or data that are streamed into the system dynamically. Today's data mining plays an important role for uncovering new trends in healthcare organizations which in turn helpful for all the parties associated with this field.

The state of the problem discussed in this research is how to predict risk of heart disease to help the doctors in health care field to predict diagnoses of this disease and suggest a suitable treatment. Decision tree algorithm was used to predict risk of heart disease.

The implementation was done using WEKA API 3.6 on (NET Beans IDE 8.1) by Java programming language on Cleveland dataset for heart disease. As a result of using decision tree algorithm (j48 classifier) the accuracy of the model is 52% for training and 79% for testing.
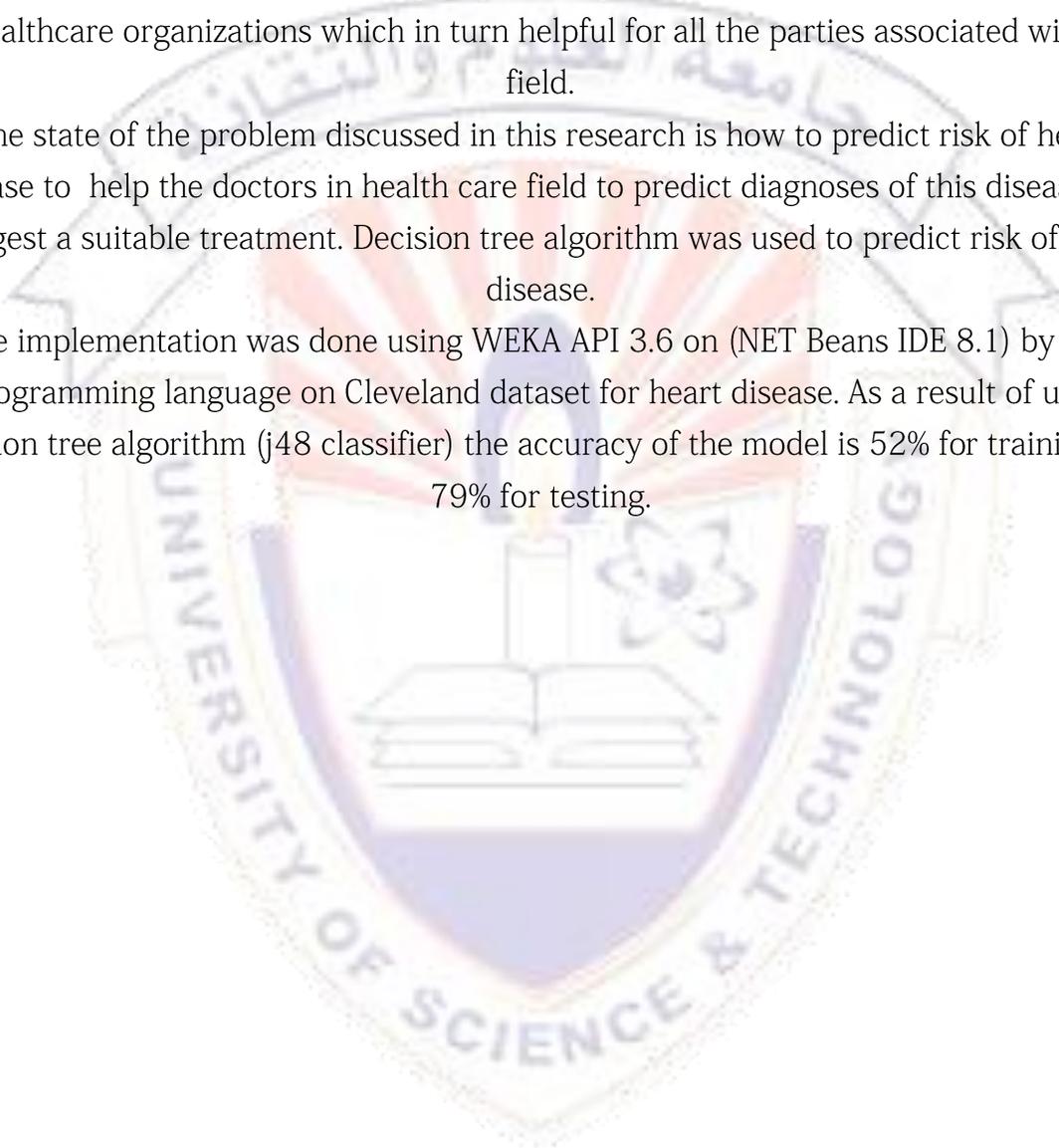
المستخلص

التنقيب عن البيانات هو عملية اكتشاف انماط مثيره للاهتمام والمعرفه من كميات هائله من البيانات .
ويمكن أن تشمل مصادر البيانات على قواعد البيانات،مستودعات البيانات،شبكة الانترنت وغيرها من
مستودعات المعلومات او البيانات التي يمكن ان تتدفق في النظام بشكل حيوي .التنقيب عن البيانات
يلعب اليوم دوراً مهما في الكشف عن الاتجاهات الجديده في مؤسسات الرعايه الصحيه التي بدورها
مفيده لجميع الاطراف المرتبطه بهذا المجال .

المشكله التي نوقشت في هذا البحث هي كيفية التنبؤ بمخاطرالاصابه بامراض القلب لمساعدة الأطباء في
مجال الرعايه الصحية للتنبؤ بتشخيص هذا المرض و تحديد العلاج المناسب. تم استخدام خوارزمية
شجرة القرار للتنبؤ بخطر الإصابه بمرض القلب.

تم تطبيق الخوارزمية عن طريق مكتبة الويكا(WEKA API 3.6) بإستخدام (NET Beans IDE 8.1)
على داتا ست متحصله من الانترنت ومفضله لدي الباحثين وهي (Cleveland data set).

ونتيجة لإستخدام خوارزمية شجرة القرار تم التوصل الى دقه للنموذج بنسبة ٥٢% للتدريب و
٧٩%للإختبار.

1.1 Research Introduction

Heart disease is major cause of morbidity and mortality in modern society. Medical diagnosis is extremely important but complicated task that should be performed accurately and efficiently. Although significant progress has be made in the diagnosis and treatment of heart disease, further investigation is still needed .The availability of huge amounts of medical data leads to the need for powerful data analysis tools to extract useful knowledge. There is a huge data available within the healthcare systems. However, there is a task off effective analysis tools to discover hidden relationships and trends in data. Researchers have long been concerned with applying statistical and data mining tools to improve data analysis on large data sets.

Disease diagnosis is one of the applications where data mining tools are proving successful results. This research proposed to find out results using data mining decision tree algorithm.

In this study we briefly examined that out of the above techniques decision tree is the most effective for the heart disease. So it is observed that, the data mining could help in the identification or the prediction of high or low risk heart disease.

1.2 Research Problem

The state of the problem discussed in this research is how to predict risk of heart disease to help the doctors in health care field to predict diagnoses of this disease and suggest a suitable treatment.

1.3 Research Objective

To implement decision tree algorithm to find out heart disease.

1.4 Research Methodology

In this research, by using data mining classification algorithm like decision tree (used j48 classifier) to classify and predict heart disease through five steps starting in prepare data set, open algorithm interface, open data set, used classifier to classify data set and output of the result, also used (NET Beans IDE 8.1) to design interfaces.

1.5 Research Structure

This research it contains five chapters. The second chapter is literature review, this chapter provides background of the data mining and its relationships to healthcare field, and other related work in this field. While the third Chapter describes methodology used in this research .after that the fourth chapter defines implementation and result for decision tree algorithm by using (J48 classifier in Weka data mining tool). While the fifth chapter is the last chapter presents conclusion and recommendations for future work.