Nile Valley University Publications
Nile Journal for Science and Engineering (NJSE)
(ISSN: ISSN :3122-5993)
Volume02, No. 01, 2026
nilevalley.edu.sd

استخدام تنقيب البيانات في التنبؤ بسرطان الثدي

دراسة حالة بمستشفى الذرة – الخرطوم

طارق عبدالكريم1 و مرشد ابراهيم2

1جامعة النيلين
2جامعة القران الكريم

المؤلف :morshedsadua@gmail.com

تاريخ الاستلام: 13 اغسطس 2025م تاريخ القبول: 17 ديسمبر 2025م

المستخلص

يقدم البحث دراسة تطبيقية عن التنقيب في البيانات واكتشاف المعرفة من البيانات الضخمة والتي غالبا ما تكون المعرفة مخفية في وسط كم هائل من البيانات "يهدف هذا البحث لاستخدام تقنيات التنقيب عن البيانات لإستكشاف المعرفة من سجلات المرضى ومعرفة اكثر الاعمار اصابة بالمرض لإجراء فحوصات وقائيةمبكرة من المرض و توفير نتائج تساهم في تقليل انتشار سرطان الثدي في الاعوام القادمة وإستخدام تقنيات تنقيب البيانات الحديثة التي تعمل علي تسهيل تحليل البيانات, تهتم الدراسة بإستخدام تقنيات تعمل علي إستخراج واكتشاف معرفة مفيدة وقابلة للاستقلال من خلال مجموعة كبيرة من البيانات . حيث نجد أن ظهور تقنيات جديدة مما ادى هذا إلي لفت الانتباه علي استخدام التقنيات في تنقيب البيانات في اكتشاف الحلول لبعض الأمراض وخاصة أمراض السرطانات حيث تضمن الدراسة في البحث في استخدام التقنيات تنقيب البيانات في سرطان الثدي باستخدام بعض الأدوات والتقنيات وتطبيق بعض خوارزميات التجميع والتصنيف باستخدام اداة التنقيب رابيدماينر وترتكز منهجية هذا البحث أولا علي تحضير البيانات التي تم الحصول عليه من مقابلة مركز الحاسوب الخاص بمستشفى الذرة ثم تطبيق تقنيات البيانات التي تم اختيارها بتسلسل معين بالرجوع إلي مجموعه من الاسباب التي يمكن تلخيصها في محورين وهما: مناسبة الطريقة لطبيعة البيانات و تلاؤمها مع أهداف البحث بالإضافة إلي كفاءة اكتشاف الأنماط مما ادى الي استخدام خوارزميات التصنيف وهي , TreeNaïve Bayes , neural networks Decisionو خوارزمية التجميع و هي K-meansلتحقيق أهداف البحث عندما تم اختيار البيانات حيث نجد أكثر الفئات العمرية عرضة للمرض هم من بين 37-46 وأكثرهم من النساء و عدد المصابين =757, حسب النوع حيث الإناث الاكثر اصابة بالمرض بعدد 6942 اصابة و الرجال الاقل بعدد 557 اصابة وان الاعوام اكثر إنتشارا للمرض حيث ان العام 2021 الاكثر انتشارا بعدد 3736 اصابة و العام الاقل 2012 بعدد 138 اصابة .

الكلمات المفتاحية : *تنقيب بيانات- البيانات الضخمة- استكشاف المعرفة - اكتشاف انماط*

The Use of Data Mining in Predicting Breast Cancer

A Case Study at Al-Durra Hospital – Khartoum

Tarig Abdelkarim[1] and Murshid Ibrahim[2]

1Neelain Univeristy

2University of Holly Guraan, Omduraman

Corresponding Author: morshedsadua@gmail.com

Received: 13th August, 2025 **Accepted: 17th Dec, 2025**

Abstract

The research presents an applied study on data mining and discovering knowledge from big data, which knowledge is often hidden in the midst of a huge amount of data.disease and provide results that contribute to reducing the spread of breast cancer in the coming years and the use of modern data mining techniques that facilitate data analysis, the study is interested in using techniques that extract and discover useful and independent knowledge through a large group of data, where we find that the emergence of new technologies. This led to drawing attention to the use of techniques in data mining in discovering solutions to some diseases, especially cancer diseases. First, I have to prepare the data obtained from the interview with the computer center of Al-Thaz Hospital Then, the application of data techniques that were selected in a specific sequence by reference to a group of reasons that can be summarized in two axes: the appropriateness of the method to the nature of the data and its compatibility with the research objectives in addition to the efficiency of pattern discovery, which led to the use of classification algorithms, namely TreeNaïve Bayes, neural networks, Decision and The aggregation algorithm, which is K-means, to achieve the objectives of the research when the data were selected, where we find that the most vulnerable age groups are among the 37-46 and most of them are women. The number of injuries was studied and distributed to the different states of Sudan. We find that the most affected states are the state of Khartoum, according to the type, where Females are the most infected with the disease, with 6942 infections, and men with the least, with 557 infections, and the years are more prevalent for the disease, as the year 2021 is the most prevalent, with 3736 infections, and the least year is 2012, with 138 infections.

Keywords: *Data mining - Big data - Knowledge exploration - Pattern discovery*