



جامعة الموصل

كلية علوم الحاسوب والرياضيات

قسم علوم الحاسوب

تنقيب قاعدة بيانات المرضى والطلاب

في كلية طب الأسنان في جامعة الموصل

رسالة مقدمة

إلى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل

كجزء من متطلبات نيل شهادة دبلوم عالٍ علوم في

علوم الحاسوب

من قبل

مروة باسم مصطفى محمد

بإشراف

أ. م. د. عمار ظاهر ياسين

## الخلاصة

تعاني كلية طب الأسنان في جامعة الموصل من عدم وجود نظام قاعدة بيانات موحدة للطلاب والمرضى المراجعين للعيادات الطبية التابعة للكلية وصعوبة توثيق الحالات المرضية وكذلك الزخم الحاصل على العيادات بسبب الاعداد الكبيرة للطلبة والتي تُعدُّ خارج الطاقة الاستيعابية للكلية (بناية الكلية صممت لاستيعاب 50 طالباً لكل مرحلة بينما اعداد طلبة المرحلة الخامسة 410 والمرحلة الرابعة 300 طالب) فضلا عن صعوبة ايجاد الحالات المرضية المناسبة للطلاب لكل فرع في العيادات التخصصية. ولهذا السبب تم هذه الرسالة تصميم وتنفيذ نظام لتتقيب بيانات الطلاب والمرضى في كلية طب الأسنان في جامعة الموصل وذلك باستخدام تطبيق إدارة قواعد البيانات Microsoft SQL Server Management Studio لتصميم وتنفيذ نظام قاعدة البيانات وبرنامج WEKA للتتقيب في قاعدة البيانات، تم اعتماد لغة Python للبرمجة واجهات النظام. الخطوات الرئيسة لقاعدة البيانات شملت التحليل والتصميم والتنفيذ كما تضمنت عملية التتقيب الخطوات السبعة، جمع البيانات، المعالجة المسبقة للبيانات، استكشاف البيانات، تحويل البيانات، نمذجة البيانات، التقييم، والنشر. انقسمت عملية تتقيب قاعدة البيانات إلى جزأين؛ الجزء الأول هو عملية عنقدة ذكية لطلاب كلية طب الأسنان للمرحلتين الرابعة والخامسة على المختبرات على عدد المقاعد المتوفرة لكل مختبر باستخدام ثلاث خوارزميات مشهورة (EM، K-Means، Canopy)، الجزء الثاني تناول تصنيف المرضى إلى اربعة أصناف حسب نوع المعالجة التي يحتاجها كل مريض باستخدام ثلاث خوارزميات مشهورة أيضا (SVM، Naïve Bayes، Random Forest) وتم استخدام مقياسي Recall و Precision لتقييم الخوارزمية الفضلى. بعد تطبيق النظام على البيانات الحقيقية لكلية طب الأسنان في جامعة الموصل تبين ان أفضل خوارزمية عنقدة هي K-Means وأفضل خوارزمية تصنيف هي Naive Bayes.

**University of Mosul**  
**College of Computer Science**  
**And Mathematics**  
**Computer Science**



**Mining Patient and Student Database System in  
the Collage of Dentistry at the University of  
Mosul**

**Marwa Bassim Mustafa**

**D.SC./Thesis**

**Computer Science**

**Supervised**

**By**

**Dr. Ammar Daher Yassin**

**2024 A.D**

**1445 A.H.**

## **Abstrac**

The College of Dentistry at the University of Mosul suffers from the lack of a unified database system for students and patients visiting the medical clinics affiliated with the college, the difficulty of documenting medical cases, and the overcrowding of the clinics due to the large numbers of students, which are considered beyond the college's capacity for each stage, while the number of students in the fifth stage is 410 (the college building was designed to accommodate 50 students and the fourth stage 300 students), in addition to the difficulty of finding appropriate medical cases for the student for each branch in the specialized clinics. This thesis includes the design and implementation of a system for mining student and patient data of the College of Dentistry at the University of Mosul, using the Microsoft SQL Server Management Studio to design and implement the database system and the WEKA program for database mining. Python language used to develop the system interfaces. The main steps of the database included analysis, design, and implementation, and the mining process included the seven steps: Data collection, data pre-processing, data exploration, data transformation, data modeling, evaluation, and dissemination. The database mining process was divided into two parts: The first part is a smart clustering process for ~~for~~ the fourth and fifth stage students at the College of Dentistry on the laboratories (on the number of chairs available for each laboratory) using three famous algorithms (Canopy, K-Means, EM). The second part dealt with classifying patients into four categories according to the type of treatment each patient needs using three well-known algorithms (SVM, Naïve Bayes, Random Forest) , Recall and Precision metrics were used to evaluate the best algorithm. After applying the system to real data from the College of Dentistry at the University of Mosul, it was found that the best clustering algorithm is K-Means and the best classification algorithm is Naive Bayes.