



ؤامعة المؤصل
كلية علوم الؤاسوب والرىاضيات

ؤؤسفن العنقءة الؤؤزفئفة بفاسؤعمال ؤوارؤمفة الفراءاعاؤ المؤضفئة المسؤوؤاة من الطبفعة

آمال مؤمود ناؤف الرؤؤوانف

رؤالة مؤاؤسؤفر
الإؤصاء

فبشراف
الأؤساء الؤكؤور
ؤكرفا فؤفف الؤؤال

المستخلص

يلقى موضوع التنقيب عن البيانات إهتماماً واسعاً في الآونة الأخيرة إذ أصبح مجالاً حيويّاً في البحث والتطوير. تعدّ العقدة من أهم الطرق المستخدمة في التنقيب عن البيانات. وتحدد عقدة البيانات مجموعة من الأنماط المتجانسة بطبيعتها في مجموعة من البيانات. تعدّ خوارزمية العقدة التجزئية K-means من أكثر الخوارزميات استعمالاً لما لها من سرعة ودقة في تحديد عدد العناقيد. إنّ مسالة الاختيار المسبق لعدد العناقيد K في هذه الخوارزمية يلعب دوراً مهماً في الحصول على نتائج دقيقة حول الظاهرة المدروسة.

تهدف الرسالة الى توظيف إحدى الخوارزميات المستوحاة من الطبيعة وهي خوارزمية اليراعات المضيئة (Firefly optimization algorithm) لتحديد العدد الأمثل للعناقيد. تم استخدام أسلوب مونت - كارلو في المحاكاة لتوليد البيانات وتقدير عدد العناقيد الأمثل. إذ أظهرت نتائج المحاكاة بالاعتماد على التباين الكلي داخل العنقود كمعيار للمقارنة تفوق الطريقة المقترحة مقارنة بطرائق اختيار عدد العناقيد الأخرى. فضلاً عن ذلك، طبقت الطريقة المقترحة على بيانات حقيقية عن الإحصاءات المالية لهيكل الجهاز المصرفي للعامين 2016 و2018 وجاءت النتائج التي حصلنا عليها مطابقة لنتائج تجارب المحاكاة، إذ ان الطريقة المقترحة فعالة للغاية ومجدية لتحديد عدد العناقيد بالاعتماد على التباين الكلي داخل العناقيد.

**UNIVERSITY OF MOSUL
COLLEGE OF COMPUTER SCIENCES
AND MATHEMATICS**



Improving partitioned clustering using Firefly algorithm nature-inspired

Amal Mahmood Naji Al Radhwani

M.Sc./Thesis

Statistics

Supervised by

Professor

Dr. Zakariya Yahya Algamal

2021 A.D.

1443 A.H.

Abstract

In recent years, data mining is now considered as one of the most vital approaches. In which, it received a wide range of interest. Clustering analysis is one of the most used data mining methods. K-means clustering has a wide use in clustering. However, the quality of K-means algorithm is fully dependent on the number of clusters, K . In this thesis, one of the nature-inspired algorithms represented by the firefly optimization algorithm is employed to determine the optimal number of clusters. A Monte Carlo study is used to show the effectiveness of the proposed algorithm over other used methods. In addition, a real data application which is related to the banking application is illustrated to show the benefits of our proposed algorithm. The simulation and real data application results reveal that the proposed method is quite efficient and feasible for determining the number of clusters in terms of the total variance between groups.