

تقييم اداء محطات مشاريع الماء باستخدام تقنية عنقدة PAM

ندوى خزل رشاد العبيدي

نعمه عبدالله الفخري

قسم نظم المعلومات الإدارية/كلية الإدارة والاقتصاد/جامعة الموصل

(قدم للنشر في ٢٠٢٢/١/١٦، قبل للنشر في ٢٠٢٢ /٤/١٣)

-الملخص:

تعتبر العنقدة واحدة من اهم الاكتشافات العلمية التي توصل اليها العلماء، إذ تتلخص الفكرة الاساسية بتجزئة البيانات الى مجموعة من العناقيد.

ويهدف البحث الى استخدام مفهوم العنقدة لنمذجة البيانات الواقعية ومدى معرفة كفاءة خوارزمية PAM في تحديد كفاءة مشاريع الماء في محافظة نينوى من ناحية الفحوصات الفيزيائية مثل درجة الحرارة والكهربائية ونسبة الحموضة اجري البحث على عينة من الفحوصات الفيزيائية مثل درجة الحرارة والكهربائية ونسبة العكور والحموضة والتي تجربها مديرية ماء نينوى في عدد من مشاريعها داخل الموصل وفي الاطراف، هذا وقد توصل البحث الى أن أداء تقنية عنقدة PAM كانت اكثر فاعلية في حساب مصفوفة المسافات للبيانات ذات الأبعاد الكبيرة، كما ادى تطبيق هذه التقنية الى انتاج قاعدة أساسية يمكن استخدامها عدة مرات وبنفس الوزن ليتم إنشاء مجال تتمحور حوله العناقيد التي تمكنا من العثور على مجاميع أكثر إثارة.

Evaluation performance of water projects stations using PAM clustering technique

Nadwa Khazal Rashad Al-Obaidi

Neema Abdullah Al-Fakhri

Department of Management Information Systems/College of Administration and Economics/University of Mosul

Abstract:

Clustering is one of the most important scientific discoveries made by scientists. The main idea is to split data into a set of clusters. The research aims to use the concept of clustering to model data and the knowledge of the efficiency of the PAM algorithm in determining the efficiency of water projects in Nineveh

Governorate in terms of physical tests such as temperature, electrical and acidity. Nineveh in a number of its projects inside Mosul and on the outskirts. The research concluded that the performance of the PAM clustering technique was more effective in calculating the distance matrix for large dimensional data, and the application of this technique led to the production of a basic rule that can be used several times with the same weight to create a field around which clusters are centered, which enables us to find more aggregates excitement.

-المقدمة :

تعتبر البيانات الركيذة المهمة لكل المشاريع والانجازات والتطور العلمي الهائل في جميع المجالات العلمية والاقتصادية والسياسية وغيرها وفي ظل ثورة المعلومات والاتصالات حيث تولدت كميات من البيانات الضخمة هذه البيانات التي نقوم نحن بتزويد شبكة الانترنت بها عبر استخدامنا الكثير والمتزايد كل يوم للانترنت من مواقع التواصل الاجتماعي واستخدام محركات البحث فان حجم البيانات سوف يزداد كثيرا فهذا الكم الهائل من البيانات ادى لزيادة الحاجة لتطوير ادوات ذكية تمتاز بالقوة لتحليل ومعالجة البيانات واستخراج المعلومات والمعارف منها فأساليب التقليدية والاحصائية لم تعد تفي بالغرض المطلوب ,من هنا ظهرت الحاجة لإيجاد طرق غير تقليدية تتعامل من خلالها مع البيانات بجميع انماطها (النصية, المرئية, الصوتية وغيرها) بحيث اصبحت البيانات متراكمة بشكل كبير في قواعد بيانات كبيرة فضلا عن تنوع المستخدم من الانسان العادي الى المتخصص فكان لابد من ايجاد طرق جديدة في محاولة لتنقيح وتبويب هذه البيانات وتصنيفها لاستخراج المعلومات وبيان مدى علاقتها ببعضها من انظمة المعلوماتية انظمة التنقيب في البيانات ضمن مخازن البيانات وهي عبارة عن تقنية تستخدم في عملية استخراج البيانات من قواعد بيانات عملاقة للإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة لصاحب البيانات. وتتم عملية التنقيب بعدة طرائق من اهمها طريقة العنقدة.

وقد تم التركيز في بحثنا هذا على استخدام طريقة العنقدة لإيجاد المراكز الابتدائية لخوارزمية، إذ تم تقسيم البيانات والتي تعمل على فصل مجموعة البيانات الى عناقيد، كل عنقود يحوي بيانات متشابهة ومتراطة داخليا تسهل التعامل معها بدلا من اختيارها بالشكل العشوائي. لتقييم صحة النتائج المستخلصة من خوارزميات العنقدة .

أولاً : منهجية البحث

١- مشكلة البحث:

يمكن بلورة مشكلة البحث بالوقوف على اهم الاسباب التي تجعل اداء بعض المشاريع افضل من الاخرى .

2- هدف البحث:

يهدف البحث الى معرفة مدة كفاءة خوارزمية PAM في تحديد كفاءة مشاريع الماء في محافظة نينوى من ناحية الفحوصات الفيزيائية مثل درجة الحرارة والكهربائية ونسبة الحموضة .

٣- اساليب البحث:

التقنية المستخدمة هي تطبيق عنقدة البيانات باستخدام خوارزمية PAM

٤- عينة البحث:

اجري البحث على عينة من الفحوصات الفيزيائية مثل درجة الحرارة والكهربائية ونسبة العكور والحموضة المعتمدة لدى مديرية ماء نينوى في عدد من مشاريعها داخل الموصل وفي الاطراف .

ثانياً- التنقيب عن البيانات:

١- تاريخ التنقيب في البيانات

ادى الانتشار الواسع لتكنولوجيا المعلومات وسهولة حصولها على زيادة في حجم المعلومات التي لم يتم ملاحظتها من قبل في التاريخ , مما يجعل التفكير في حجم هذه البيانات موضوع الجدل من حيث جدوى وجودة بهذه الطريقة العشوائية. ومن هنا ظهر مصطلح اكتشاف المعرفة في اواخر الثمانينات ولم يصاغ مصطلح التنقيب عن البيانات حتى التسعينات كتقنية لاكتشاف المعرفة ولتحليل الكميات الهائلة من البيانات وتحويلها من مجرد بيانات متراكمة وغير مفهومة الى بيانات ذات قيمة يمكن فيما بعد استخدامها لقد حظيت مرحلة التنقيب عن البيانات باهتمام كبير في مجال البحث العلمي في العقد الماضي , بهدف تطوير و بناء الخوارزميات الرياضية التي تعتبر هي اساس التنقيب عن البيانات وهي مشتقة من العديد

من التخصصات العلمية: وهي الذكاء الاصطناعي (ذكاء شبيه بالبشر تعرضه الآلات) والاحصاء والرياضيات والمنطق (الدراسات العددية لعلاقات البيانات) وعلوم التعلم والتعرف على الانماط وتعلم الآلة (الخوارزميات التي تتعلم من البيانات لعمل التنبؤات) . وغيرها من العلوم الذكية وغير التقليدية [١١].

٢- مفهوم التنقيب عن البيانات

التنقيب عن البيانات او ما يسمى اكتشاف المعرفة هو عملية تحليل البيانات الضخمة المتعددة الانواع والمصادر للبحث عن وجود أي انماط او علاقات مختلفة بين هذه البيانات الضخمة وتحويل هذه الاكتشافات الى معلومات هو عملية الكشف والعثور على معلومات ذات فائدة من خلال استعمال البيانات ويشار الى عملية التنقيب في البيانات على انها عملية احصائية ثنوية. وتجدر الاشارة الى ان جذور تقنيات التنقيب عن البيانات ترجع الى الاساليب الاحصائية في التحليل فعند التعامل مع حجم كبير من البيانات تظهر مسائل جديدة مثل كيفية تحديد النقاط المميزة في البيانات وكيفية تحليل البيانات في فترة زمنية معقولة وكيفية قرار ما اذا كانت أي علاقة ظاهرية تعكس حقيقة في طبيعة البيانات حيث يكون الغرض عادة تعميم النتائج على كامل البيانات ليكون الهدف من تنقيب البيانات هو اختزال كميات البيانات الكبيرة او ضغطها بحيث تعبر عنها بشكل بسيط عن كامل البيانات بدون تعميم . أي انه عملية استخراج انماط او معلومات مفيدة من مجموعة من البيانات عن طريق الاحصاء والذكاء الاصطناعي وعن طريق قواعد البيانات . وهو عملية الكشف على معلومات من خلال استعمال مجموعة من الأدوات المعقدة. بعض من هذه الأدوات تشمل أدوات الإحصاء الاعتيادية والرسوم البيانية الحاسوبية[٢]. فالتنقيب في البيانات منهجية تجمع بين نتائج الأبحاث في الذكاء الاصطناعي، الفهم الآلي، التعرف على الأشكال، قواعد المعلومات، الرياضيات الإحصائية، واجهات الاستعمال واللغة. وبحسب مراكز الأبحاث يمكن الفصل بين نوعين من استخراج المعلومات فتسمية (KDD Databases Mining in Discovery Knowledge) تستعمل عند الباحثين في الذكاء الاصطناعي والفهم الآلي وتسمية Data Mining تستعمل عند الباحثين في الرياضيات الإحصائية أو خبراء المعلومات [٣] [٥].

٣- مراحل اكتشاف التنقيب عن البيانات:

لا يعد اكتشاف المعرفة في قاعدة البيانات عملية سهلة, إذ انها لا تنتهي بجمع البيانات فقط, بل يتم استخلاص المعلومات المفيدة من قواعد البيانات او من مجموعة البيانات ويتم ذلك عن طريق المراحل التالية [١٣] [٤]:

١. اكتشاف البيانات : تتضمن مرحلة جمع البيانات واكتشاف وتحديد وتوصيف البيانات المتاحة.

٢. تصفية البيانات وتنقيتها : تحذف البيانات التي ليس لها اهمية والغير متناسقة والمتضاربة والغير ضرورية في قواعد البيانات.
٣. تكامل البيانات: توحيد شكل البيانات اي يتم جمع البيانات المتشابهة وذات الصلة وتدمج مع بعضها
٤. اختيار البيانات : تعني تحديد واسترجاع البيانات الملائمة من قاعدة البيانات
٥. تحويل البيانات : تحويل البيانات الى نماذج مخصصة ملائمة لإجراء البحث والدراسة
٦. التنقيب عن البيانات : يتم استخدام الطرق الذكية لاستخراج الانماط من البيانات المتوفرة لدينا.
٧. تقييم النمط: يحدد الانماط التي نحتاجها للمشكلة المطروحة ونبدأ بفصلها لوحدها
٨. عرض وتقييم المعرفة : وهي عرض النتائج حيث يتم عرض الانماط او البيانات والمعلومات المفيدة التي تم التوصل اليها بطريقة مرئية وعلى هيئة رسومات بيانية لمساعدة المستخدم فهم وتفسير نتائج استخراج البيانات

٤- مفهوم العنقدة :

عنقدة البيانات هي عملية وضع البيانات في مجموعات متماثلة. خوارزمية العنقدة تقسم مجموعة من البيانات إلى عدة مجموعات. حيث أن التشابه بين النقاط ضمن مجموعة معينة أكبر من التشابه بين نقطتين ضمن مجموعتين مختلفتين. فكره تجمع البيانات هي تعد العنقدة واحدا من نتائج التطور العلمي في مجال تصنيف البيانات وتبويبها , فهي من التقنيات الشائعة والمفيدة في اكتشاف المجاميع المتعددة وتعريف الأنماط والتوزيعات المهمة في البيانات الخام. وتعود بدايات العنقدة إلى النصف الأول من القرن الماضي اذ وضع أسسها الأولى العالم البريطاني. Ronard fisher إن الفكرة الأساسية للعنقدة تتلخص بتجزئة مجموعة البيانات إلى عنقايد , اذ إن نقاط البيانات للعنقود الواحد تتشابه أكثر مع بعضها البعض مقارنة مع تلك النقاط في العناقيد الأخرى K وقد كانت طرائق التصنيف سابقاً تعتمد على خبرة الخبراء وعلى الجهد الشخصي الكبير في هذا المجال لتؤدي إلى نتائج دقيقة, ولكن التوسع العلمي الكبير في مجال الحاسوب جعل عملية السيطرة على المعلومات الكبيرة أمر سهل وبكلفة وجهد اقل مما جعل العلماء يستغلون الإمكانيات الجبارة لهذا الجهاز العظيم لمصلحة اختصاصاتهم . إن العنقدة يمكن تعريفها بأنها عبارة عن عملية تحليل إحصائي متعدد المتغيرات

ومبرمج ويتم من خلاله دراسة التشابه بين مجموعة من المتغيرات لنماذج مختلفة ومتعددة ثم مقارنة هذه النماذج مع بعضها البعض بالاعتماد على ما تحتويه من متغيرات وترتيب العلاقة مع بعضها بشكل عناقيد، فيمكننا إيجاد التشابه بين المتغيرات المختلفة بعد إجراء المقارنة وترتيبها بشكل عناقيد بالاعتماد على مدى تواجدها في النماذج المختلفة، ويمكن أن يطلق على العنقدة مصطلحات عديدة حسب التخصص والمجال الذي تدرس فيه فمثلاً تسمى بالتجزئة في نظرية البيانات، التعلم غير الموجه في تمييز الأنماط، علم التصنيف العددي في علوم الحياة والبيئة، وتستخدم خوارزميات التجميع على نطاق واسع ليس فقط لتنظيم وتصنيف البيانات وإنما هي مفيدة لضغط البيانات وبناء نموذج مرتب لها، وبالتالي نتمكن من بناء نموذج للمشكلة على أساس تلك المجموعات [١] [٦].

٥-انواع العنقدة:

يمكن تصنيف خوارزميات العنقدة تبعاً لنوع المتغيرات الموجودة في البيانات وكمايلي [١٠]:
تتصنف العنقدة بناءً على المعيار او المقياس الذي يعرف مدى التشابه بين عناصر البيانات المدخلة الى الخوارزمية، وكذلك على المفاهيم النظرية والاساسية التي تعتمد عليها تقنية تحليل العنقدة مثال على ذلك النظرية المضببة والاحصاء .
ولقد تطورت خوارزميات العنقدة بناءً على المواضيع المتعلقة بها مثل تعدين البيانات، والتعلم الآلي، وتمييز الأنماط،... وغيرها، وصُنفت تلك الخوارزميات إلى الأصناف الآتية:

١ . العنقدة الهرمية

٢ . العنقدة التجزئية.

٣ . عنقدة البيانات ذات الابعاد الكبيرة.

٤ . العنقدة مع الصفات الوراثية.

وصنف بعض الباحثين خوارزميات العنقدة طبقاً للطريقة المعطاة في تعريف العناقيد الى الاصناف

الآتية:

١ . العنقدة الهرمية.

٢. العنقدة التجزئية.

٣. العنقدة المعتمدة على الكثافة.

٤. العنقدة المعتمدة على الشبكة المتعامدة.

ولكل صنف من الأصناف المذكورة أعلاه أنواع متفرعة عنها وخوارزميات مختلفة لإيجاد العناقيد، وعليه فيمكن تصنيف خوارزميات العنقدة تبعاً لنوع المتغيرات الموجودة في البيانات وعلى النحو الآتي [17][15]:

١. العنقدة الإحصائية: ويعتمد هذا النوع على مفاهيم التحليل الإحصائي، فيعتمد الى مقاييس التشابه لتجزئة النماذج التي تكون محددة ببيانات عديدة.

٢. العنقدة الإدراكية : تُستخدم لعنقدة البيانات الصنفية ، حيث تُعقد النماذج تبعاً للمفاهيم التي تحملها وليس لقيمها.

٣. العنقدة المضطربة: تستخدم تقنيات مضطربة لبيانات العنقود، إذ يفترض انه يمكن تصنيف النماذج إلى أكثر من عنقود واحد. وتستخدم البيانات الحقيقية غير المؤكدة (غير الموثوقة).

٤. العنقدة الهشة: يفترض هذا النوع تجزئة غير متداخلة ، وهذا يعني ان نقطة البيانات إما أن تنتمي إلى العنقود أو لا تنتمي.

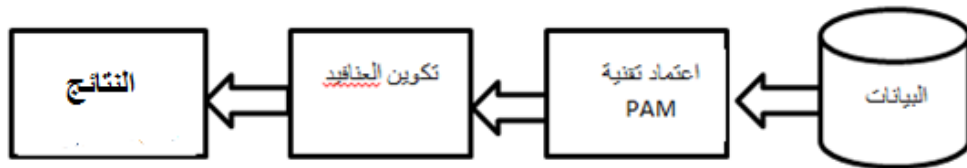
٥. عنقدة شبكة كوهين : و تعتمد العنقدة هنا على مفاهيم الشبكة العصبية، ويتشكل العنقود من العقد الخارجية.

ثالثاً- الجانب العملي:

الماء هو سر الحياة، وهو الهدف السادس من اهداف التنمية المستدامة الذي يؤكد على توفير المياه النظيفة والصالحة للشرب للجميع وهو ايضاً حق من حقوق الانسان. وقبل استخدام مياه الشرب يجب معالجتها، ويفضل تقدم تقنيات الذكاء الاصطناعي أصبح من الممكن إجراءات الرقابة على محطات معالجة المياه بوسائل أرخص وأسهل مما يساعد بدوره في حماية الصحة العامة.

أجري البحث على البيانات المستحصلة من مديرية ماء نينوى للفحوصات الفيزيائية التي تقوم بإجرائها بها كوادرات المختبرات التابعة للمديرية بشكل دوري، تضمنت الفحوصات مقاييس الكهربائية والـ PH و درجات الحرارة و الـ Turbidity وهو مقياس العكورة. أخذت العينات لشهري كانون الثاني وشباط للنصف

الاول من السنة (٢٠١٩) كون الامطار شديدة في هذين الشهرين ومن الممكن حدوث سيول وانهيارات طينية، كذلك شهري حزيران وتموز (٢٠١٩) كون الماء يكون نقي صافي واجراء المقارنة بين اداء المشاريع لهذين الفصلين من السنة، والملحق(١) يوضح البيانات المستخدمة في البحث. ينقسم الجانب العملي الى قسمين، الاول هو اعتماد تقنية PAM اي التجزئة حول الـ medoid ، والثاني مقارنة اداء المشاريع بين النصف الاول والنصف الثاني من السنة، والشكل(١) يوضح مخطط للجانب العملي، وقد استخدم برنامج Matlab2016 لتنفيذ مراحل الجانب العملي.



الشكل (١)

مخطط الجانب العملي من البحث

-تطبيق تقنية عنقدة PAM على البيانات-

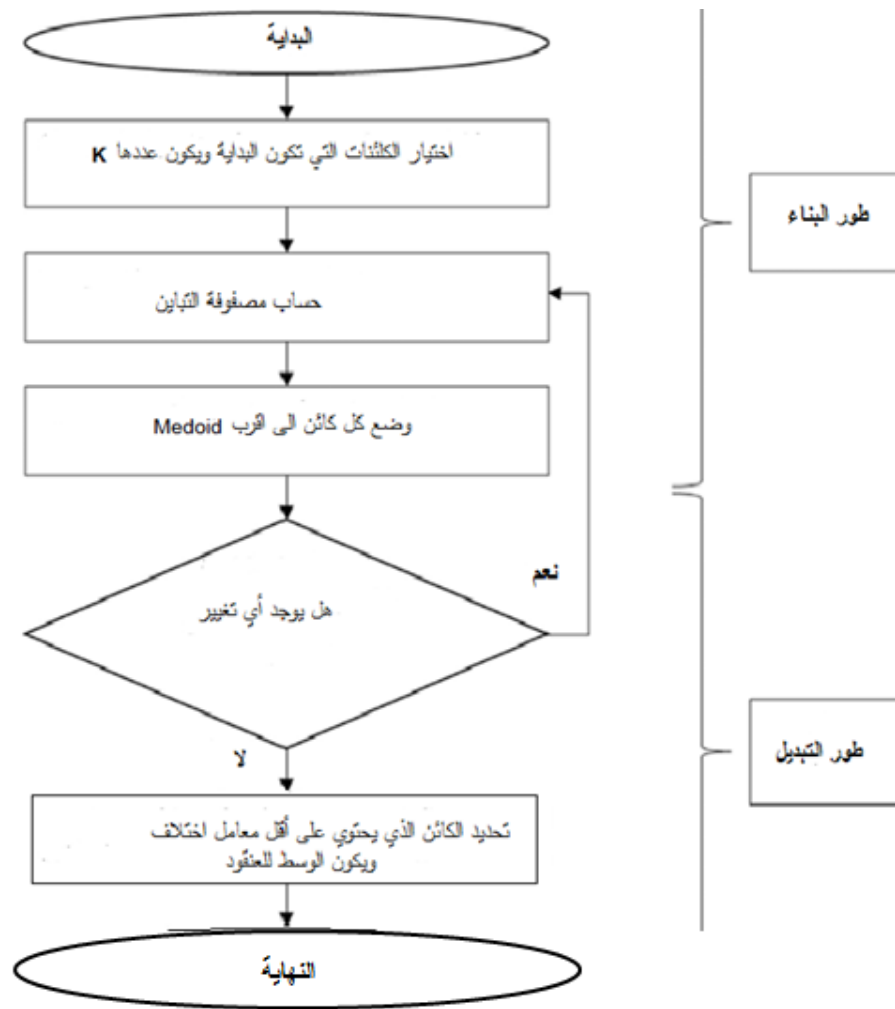
١- تقنية عنقدة PAM:

المصطلح PAM هو اختصار لجملة Partitioning Around Medoid أي التجزئة حول الوسط أو المركز وتعمل PAM مثل جميع تقنيات العنقدة بطريقة تكرارية إذ يتم اختيار العناقيد والصفات بشكل عشوائي والمرونة في استبدال العناصر ذات صفات معينة بصفات أخرى هو مقياس لجودة العنقدة وتستمر عملية التقسيم الى حد لا يمكن فيه ان تؤدي عملية التقسيم الى اي تغيير او تحسن في العناقيد المتكونة [12]. يؤدي التقسيم إلى تجميع الكائنات حول k-medoids ، و يتم تحديد k مسبقاً. تأخذ الخوارزمية شكل صعود التل شديد الانحدار، تكون العناقيد بداية فارغة ويتم اختيار الكائن الذي ينتج أفضل تجميع وبعملية مبادلة بسيطة لكل كائن متكرر او متقارب حول الوسط تتم اضافة الكائنات الواحد تلو الآخر ويتم تحديد k-medoids ويتم تحديد الوسط الجديد في كل خطوة للتقليل من دالة الهدف المستخدمة [14][9].

٢- مراحل عمل الخوارزمية [7]:

١. التهيئة: تحديد k من نقاط البيانات عشوائياً و n كمتوسطات

٢. خطوة التعيين: اربط كل نقطة بيانات بأقرب medoid.
 ٣. خطوة التحديث: تتم لكل medoid وكل نقطة بيانات وحساب التكلفة الإجمالية (أي متوسط الاختلاف لجميع نقاط البيانات).
 ٤. تكرار عملية الابدال في ٢ و ٣ الى حين الوصول الى المرحلة التي لا يكون هناك تغيير في القيم. والعنقدة بـ PAM تمر بطورين، الطور الاول ويسمى بطور البناء ويتم فيه اختيار الكائنات البدائية ويكون اختيارها عشوائياً لتكون هي الوسط، وحساب مصفوفة التباين لتجميع الكائنات الى اقرب medoid ، اما الطور الثاني فهو طور التبديل (الابدال) فباستخدام عملية مبادلة بسيطة. لكل كائن ميدوي متكرر i وكائن غير ميداني j يتم تحديدهما للحصول على أفضل عنقود [٨].
- خوارزمية k-medoids والتي هي اساس عمل PAM فهي خوارزمية عنقدة مرتبطة بخوارزمية k-mean و medoidshift. وكلتا الخوارزميتين تقسم مجاميع البيانات إلى مجموعات تجزئية (k) في محاولة لتقليل إجمالي الخطأ التربيعي ، بينما تقلل k-medoids من مجموع الاختلافات بين النقاط المصنفة ونضعها في مجموعة ونقطة معينة كمركز لتلك المجموعة (العنقود) ويتم اختيار النقاط التي تكون مراكز البيانات من قبل k-medoids [16]



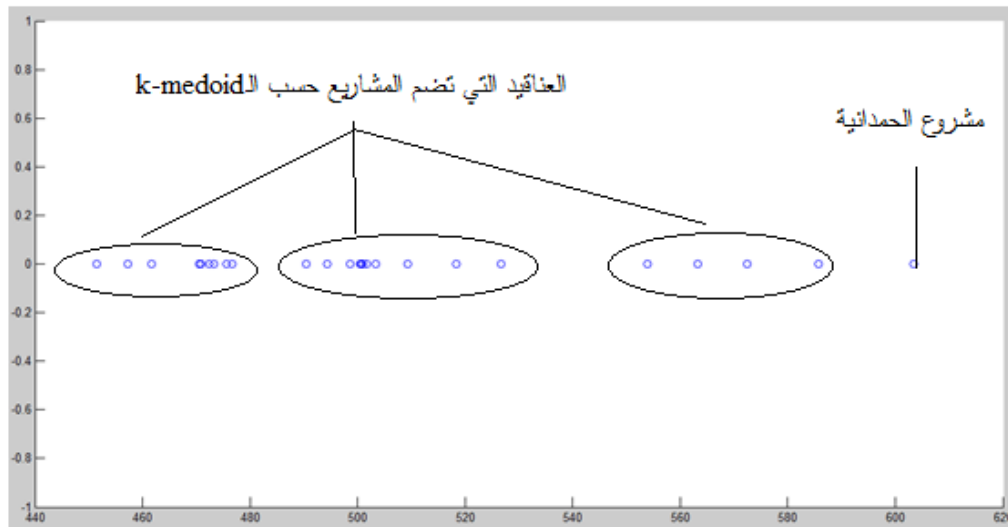
الشكل (٢)

توضيح الاطوار التي تمر بها العنقدة بتقنية PAM

بداية نُفذت الخوارزمية على البيانات الخاصة بشهر كانون الثاني من العام ٢٠١٩، إذ عمدت الخوارزمية الى تجميع المشاريع ذات الفحوصات المتقاربة ووضعها في عناقيد ، فتولد لدينا ثلاثة عناقيد وكما يوضحها كل من الجدول (١) والشكل (٣) باستثناء مشروع ماء الحمدانية الذي كانت نتائجه عالية فقد شدّ عن البقية، مع ملاحظة ان النتائج، كلما كانت قليلة كلما كان ذلك افضل.

الجدول (١): مشاريع الماء مرتبة تصاعديا بعد تنفيذ PAM لشهر ٢٠١٩/(١)

ت	المشروع	K-medoid
1.	الأيمن الموحد/المنتج	451.5
2.	الأيمن الموحد/المنتج	457.2
3.	خواجة خليل/المنتج	461.7
4.	خواجة خليل/المنتج	470.5
5.	الأيمن الجديد/المنتج	470.8
6.	القية/المنتج	472.44
7.	الأيمن الموحد/المنتج	473.2
8.	الايسر الجديد/المنتج	475.65
9.	النمرود/المنتج	476.7
10.	بيسان/المنتج	490.36
11.	الدندان/المنتج	494.3
12.	التوسيع/المنتج	498.6
13.	الحدباء/المنتج	500.4
14.	الدندان/المنتج	500.64
15.	برطلة/منتج	500.95
16.	الزهور/المنتج	501.73
17.	الزهور/المنتج	503.35
18.	الايسر القديم/المنتج	503.4
19.	الساھرون/المنتج	509.3
20.	الساھرون/المنتج	518.3
21.	الحمداية/منتج	526.6
22.	الساھرون/المنتج	553.9
23.	بعشيقه/منتج	563.3
24.	السلامية/المنتج	572.5
25.	بعشيقه/المنتج	585.72
26.	الحمداية/منتج	603.4

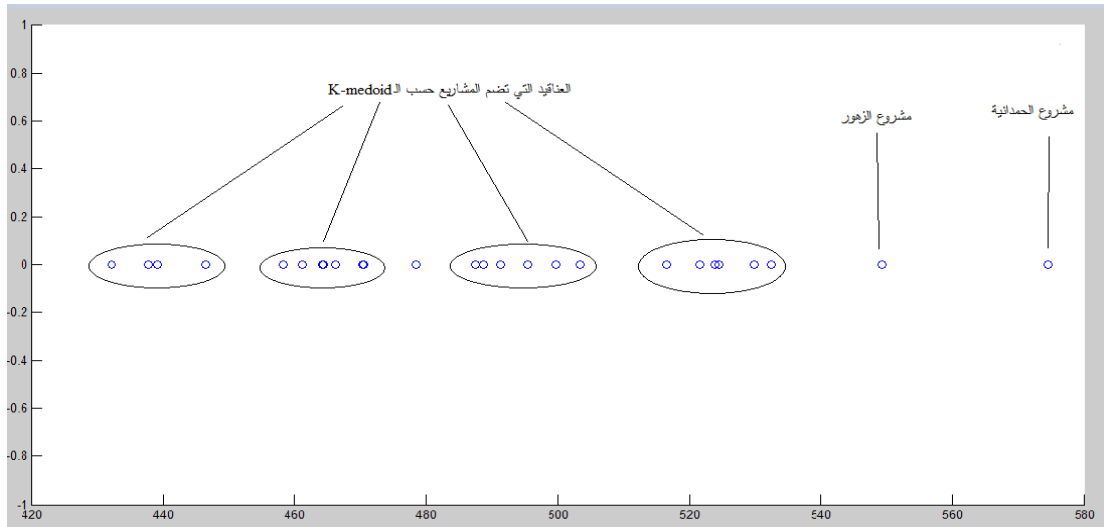


الشكل (٣): نتائج تنفيذ PAM على بيانات شهر كانون الاول/٢٠١٩

يوضح الشكل (٣) ان المشاريع ذات المقاييس (الفحوصات) المتقاربة في النتائج حسب ما موضح في الجدول (١) قد تجميعها في ثلاثة عناقيد العنقود الاول من اليسار الى اليمين ضم المشاريع الايمن الموحد والايمن الجديد وخواجة خليل والقبة مع ملاحظة ان هناك اكثر من قراءة لبعض المشاريع وكانت هذه افضل المشاريع نظرا لقربها من النهر اضافة الى اجهزتها ومعداتها المتطورة ، العنقود الثاني في الوسط وضم مشاريع الايسر الجديد، والنمرود، وبيسان، والقبة، والحدياء، والدندان، والزهور، اما العنقود الثالث فقد ضم الايسر القديم، والسلامية، والساهاون، وبعشيقه، اما مشروع ماء الحمدانية في القراءة الثانية فقد شذ عن مستوى بقية الفحوصات. والان نأخذ نتائج تنفيذ عنقده PAM لشهر شباط كما يقرأها الجدول (٢) ويوضحها الشكل (٤).

الجدول (٢): مشاريع الماء مرتبة تصاعديا بعد تنفيذ PAM لشهر (٢)/٢٠١٩

ت	المشروع	K-medoid
1.	الأيمن الموحد /المنتج	432.2
2.	الزهور /منتج	437.7
3.	بيسان	439.15
4.	الساهاون /منتج	446.6
5.	الأيمن الموحد /المنتج	458.2
6.	التوسيع	461.2
7.	الدندان /منتج	464.2
8.	السلامية /منتج	464.4
9.	الساهاون /منتج	466.2
10.	الايسر الجديد /منتج	470.3
11.	الأيمن الموحد /المنتج	470.4
12.	بعشيقه	478.45
13.	الزهور	487.5
14.	خواجة خليل /منتج	488.8
15.	بعشيقه /منتج	491.23
16.	الحمدانية /منتج	495.45
17.	النمرود /منتج	499.7
18.	الأيمن الجديد /منتج	503.4
19.	الدندان /منتج	516.5
20.	الحدياء	521.5
21.	خواجة خليل /منتج	523.8
22.	القبة	524.4
23.	الايسر القديم /منتج	529.8
24.	برطلة /منتج	532.4
25.	الحمدانية /منتج	549.3
26.	الزهور /منتج	574.5



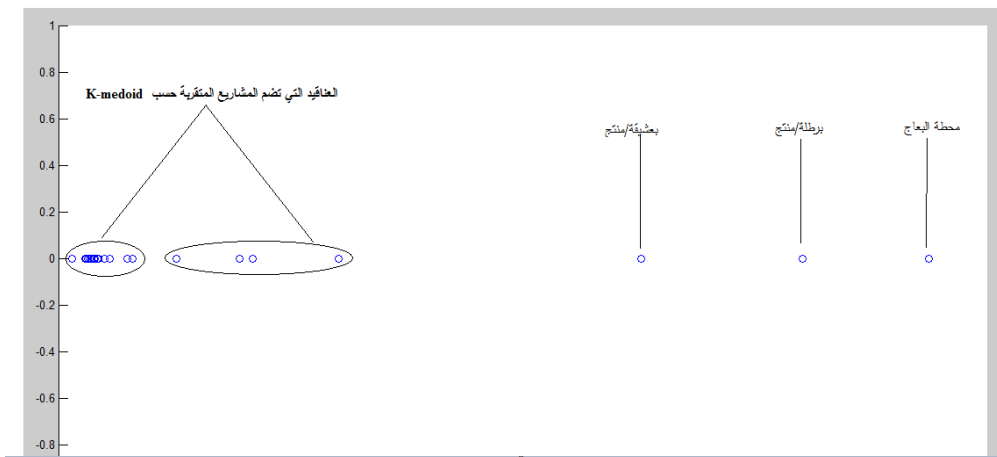
الشكل (٤): نتائج تنفيذ PAM على بيانات شهر شباط/٢٠١٩

يوضح الجدول (٢) والشكل (٤) ان نتائج شهر شباط كانت بمجملها افضل من نتائج الشهر الاول وذلك لان كمية الامطار المتساقطة كانت اقل منها حدة من الامطار المتساقطة على المدينة للشهر الاول.

اما نتائج تنفيذ عنقدة PAM لفصل الصيف فتظهرها الجداول (٣) و(٤) والاشكال (٥) و (٦)

الجدول (٣): ترتيب المشاريع بعد تنفيذ PAM لشهر ٢٠١٩/٦

K-medoid	اسم المشروع	ت
406.45	خواجة خليل	.1
413.1	القديم /منتج	.2
413.3	خواجة /خليل /منتج	.3
413.3	الايبر الجديد /منتج	.4
413.5	خواجة خليل /منتج	.5
414.1	الأيمن الموحد /منتج	.6
415	القبية /منتج	.7
416.1	الحدباء /المنتج	.8
417.15	التوسيع /منتج	.9
417.8	بيسان /المنتج	.10
419.1	التوسيع /منتج	.11
419.17	الحدباء /المنتج	.12
422.3	الزهور /منتج	.13
433.4	القيارة /منتج	.14
435.75	الساهاون /منتج	.15
457.1	الحمدانية /المنتج	.16
487.8	الدندان /منتج	.17
536	برطلة	.18
682.6	بعشيقه /المنتج + ابار	.19
760.8	برطلة /منتج	.20
822.26	محطة البعاج / ابار	.21

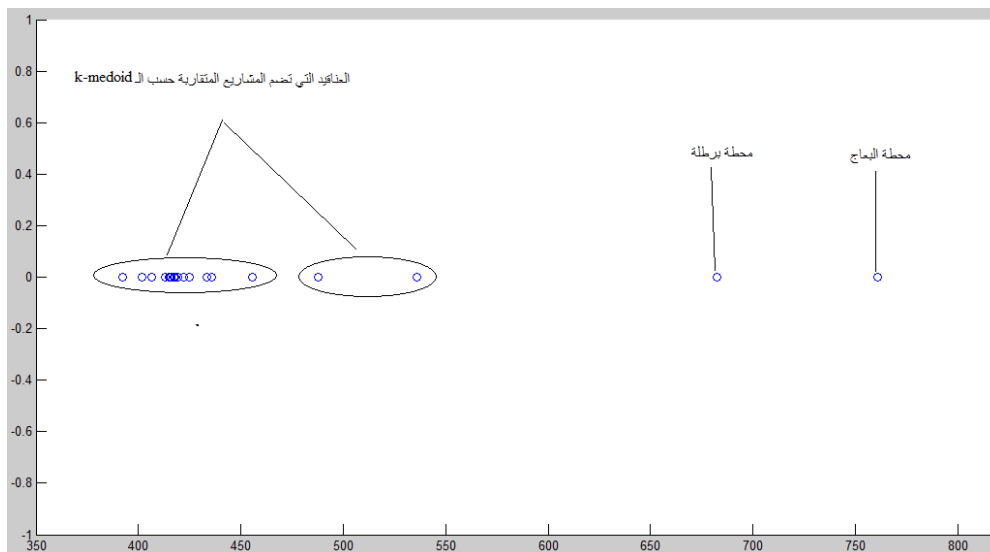


الشكل(٥):نتائج تنفيذ PAM على بيانات شهر حزيران/٢٠١٩

تظهر النتائج التي حصلنا عليها في الجدول (٣) ان هناك تقارباً وهذا يؤكد الشكل (٥) باستثناء محطات بعشيقه و برطلة والبعاج كانت نتائجها شاذة عن البقية اما نتائج شهر تموز فيوضحها كل من الجدول (٤) والشكل (٦).

الجدول (٤): ترتيب المشاريع بعد تنفيذ PAM لشهر تموز/٢٠١٩

K-medoid	اسم المشروع	ت
392.3	النمرود /منتج	.1
401.65	خواجة /خليل /منتج	.2
406.45	خواجة خليل المنتج	.3
413.1	الايسر القديم /منتج	.4
413.3	الايسر الجديد /منتج	.5
415	القبة /منتج	.6
415.25	الأيمن الجديد /منتج	.7
416.1	الحدباء /المنتج	.8
417.15	التوسيع /منتج	.9
417.15	الأيمن الموحد /منتج	.10
417.8	بيسان /المنتج	.11
419.1	التوسيع /منتج	.12
419.17	الحدباء /المنتج	.13
422.3	المنتج	.14
425	القيارة /منتج	.15
433.4	القيارة /منتج	.16
435.75	الساهاون /منتج	.17
455.8	الحمدانية /المنتج	.18
487.8	الذندان /منتج	.19
536	برطلة	.20
682.6	بعثيقة /المنتج + ابار	.21
760.8	برطلة /منتج	.22
822.26	محطة البعاج / ابار	.23



الشكل (٦): نتائج تنفيذ PAM على بيانات شهر تموز/٢٠١٩

-تفسير النتائج:

أظهرت نتائج الجانب العملي ان الفحوصات التي أجريت على عينات الماء أفرزت ثلاثة عناقيد في شهر كانون الثاني واربعة عناقيد في شهر شباط لان سنة ٢٠١٩ كانت سنة مائية فاختلفت النتائج نتيجة لكثرة الامطار والمصحوبة بانجرافات التربة وكذلك طبيعة المنطقة الجغرافية حتى اصبحت فلاتر المشاريع المائية لا تستوعب كميات المياه الواردة اليها، وكان مشروع ماء الحمدانية هو الاكثر بعداً عن المشاريع الاخرى لان المياه التي تصل الى هذا المشروع تمر ثلاث مراحل وصولاً اليه، وهذا ما توضحه الجداول ١ و ٢ وكذلك الاشكال ٣ و ٤.

أما نتائج شهري حزيران وتموز فقد أفرزت لنا عنقودين فقط وكانت نتائجهم متقاربة لان في هذين الشهرين من السنة يكون الماء انقى ما يمكن لعدم وجود الامطار وكذلك فان الشوائب في ماء النهر قد ترسبت مع جريان النهر، اما المشاريع البعيدة عن المشاريع الاخرى فانها تأخذ المياه من الابار والمياه الجوفية مثل مشروع ماء البعاج اضافة الى ما يصلها من ماء النهر، وهذا توضحه الجداول ٣ و ٤ والاشكال ٥ و ٦.

-الاستنتاجات:

- ١- أظهرت نتائج البحث أن أداء PAM كان اكثر فاعلية في حساب مصفوفة المسافات للبيانات ذات الأبعاد الكبيرة أما في حالة البيانات ذات الأبعاد الصغيرة فإن PAM تأخذ جزءاً أكبر من وقت التشغيل .
- ٢- أدت خوارزميات التجميع القائمة على الـ medoid والمستخدمه في الجانب العملي إلى تحسين التجميع (العناقيد) والمرونة الإلكترونية. كما أن تطبيق خوارزميات العنقدة الهرمية على تجميع القواعد يكشف أيضاً عن معلومات مفيدة.
- ٣- أدى تطبيق تقنية PAM الى انتاج القاعدة الأساسية التي تستخدم عدة مرات وبنفس الوزن ليتم إنشاء الحيز الذي تتمحور حوله العناقيد التي تمكننا من العثور على مجاميع أكثر إثارة للاهتمام والوقوف عندها لدراستها.
- ٤- كلما كان الموقع قريباً من النهر او مصادر المياه تكون النتائج اكثر وضوحاً ومصداقية.

٥- إن المناطق التي تعتمد على مياه الابار او المياه الجوفية تكون مقاييسها الفيزيائية عالية نظرا لمرور المياه بعدة مراحل قبل وصوله للفحص.

-المقترحات:

١- في حال أردنا تحسين تحليل البيانات من خلال عنقدة البيانات والقواعد المرتبطة بها، فمن الافضل استخدام طرق اخرى لتلخيص المعلومات التي تقدمها العناقيد ومقارنتها مع هذه الطريقة.

-المصادر:

أ-المصادر باللغة العربية:

١- ابو اليمين، عمرو عادل حمزة (٢٠١٩)، "استخدام تقنية تنقيب البيانات لتحليل بيانات المستثمرين بمفوضية تشجيع الاستثمار"، رسالة ماجستير، ص ص ٢١-٢٢، كلية الدراسات العليا، جامعة النيلين، الخرطوم-السودان.

٢- الفخري، نعمه عبدالله (٢٠١٩)، " أساسيات في تعدين البيانات مفاهيم وتقنيات"، ص٣٦، ط١، دار الايام للنشر والتوزيع، عمان-الاردن.

٣- المبارك، البدوي سعد البدوي(٢٠١٧)، "استخدام تقنيات تنقيب البيانات لاستكشاف انماط مؤثرات التحصيل الاكاديمي لطلاب المرحلة الثانوية-(دراسة حالة شهادة التعليم الثانوي الصف الثالث علمي من العام ٢٠٠٧ حتى العام ٢٠١٦)"، رسالة ماجستير، ص ص ٣٢-٣٣، كلية علوم الحاسوب وتقانة المعلومات، جامعة النيلين، الخرطوم-السودان.

٤- سيد، احمد فايز احمد (٢٠١٦)، " أدوات التنقيب عن البيانات مفتوحة المصدر-دراسة تحليلية تقييمية" مجلة جامعة طيبة للاداب والعلوم الانسانية، السنة الخامسة، العدد ١٠، ص ص ٧٩٦-٧٩٧، جامعة قناة السويس-مصر.

٥- علي، هدى عبدالرحيم حسين (٢٠١٨)، "استخدام تقنية التنقيب عن البيانات لتحليل المؤشرات المالية لعينة من المصارف الاهلية العراقية باعتماد خوارزمية CART"، المجلة العراقية لتكنولوجيا المعلومات، المجلد ٩، العدد ٢، ص ص ٣٦-٣٧.

٦-يوسف، ايمن و محمودي، حاتم و علي، تغريد(٢٠١٧)، "توظيف التنقيب في البيانات لدعم الصيانة التنبؤية باستخدام الحاسب"، مجلة جامعة طرطوس للبحوث والدراسات العلمية،-سلسلة العلوم الهندسية، المجلد (١)، العدد(١)، ص ١٤١، سوريا.

ب-المصادر الاجنبية:

- ٧-A D & Indahwati & Sartono B. Alfarizy,(2017), " **Clustering box office movie with Partition Around Medoids (PAM) Algorithm based on Text Mining of Indonesian subtitle**", IOP Conference Series: Earth and Environmental Science, PP. 3-4
- ٨-Al Abid Faisal Bin (2014), "A Novel Approach for PAM Clustering Method", International Journal of Computer Applications (0975 – 8887) Volume 86 – No 17, P.2.
- ٩-Budijai ,Weksi & Friedrich, Leisch (2019), " **Simple K-Medoids Partitioning Algorithm for Mixed**", www.mdpi.com/journal/algorithms, PP. 2-3.
- ١٠-Faizan ,Muhammad & Zuhairi, Megat F. & Ismail ,Shahrinaz & Sultan,Sara (2020) "Applications of Clustering Techniques in Data Mining: A Comparative Study" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 12, P.148, India.
- 1١-Hong,Xiaobin (2021) "Application of Data Mining Technology in Software Engineering", Journal of Physics: Conference Series, 2066 (2021) 012013, IOP Publishing, P.2m Canada.
- 1٢-K. , Chitra,& Maheswari ,D (2017), " A Comparative Study of Various Clustering Algorithms in Data Mining". IJCSMC, Vol. 6, Issue. 8, PP.110 – 111.
- 1٣-Mughal, Muhammd Jawad Hamid (2018), "Data Mining: Web Data Mining Techniques, Tools and Algorithms: An Overview", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 6, PP 209-210, Dubai-United Arab Emirates.
- 1٤-Rechkalov,Timofey V.(2016), "Partition Around Medoids Clustering on the Intel Xeon Phi Many-Core Coprocessor", Ministry of education and science of the Russian Federation "Research and development on priority directions of scientific-technological complex of Russia for 2014–2020" Federal Program, PP.30-31.
- 1٥-Saxena , Amit & Prasad, Mukesh & Gupta, Akshansh & Patel, Om Prakash & Bharill, Neha & Tiwari, Aruna & Er, Meng Joo & Ding, Weiping & Lin, Chin-



- Teng (2017) "**A Review of Clustering Techniques and Developments**", Springer , Neurocomputing, Volume 267, Pages 664-681
- 1٦-Schubert, Erich & . Rousseeuw, Peter J,(2019), "**Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms**", arXiv:1810.05691v3 [cs.LG] .
- 1٧-Son, Le Hoang & Hai, Pham Van (2016), "**A Novel Multiple Fuzzy Clustering Method Based on Internal Clustering Validation Measures with Gradient Descent**", International Journal of Fuzzy Systems volume 18,PP.894–903.

الملحق (١)
البيانات المستخدمة بالبحث (شهر ٢٠١٩/١)

E.C	PH	Tem	Tur	المشروع
546	7.8	18	49	الساهاون /خام
535	7.7	18	4.3	المنتج
446	7.9	18	48	الأيمن الموحد /الخام
432	7.8	18	3.8	المنتج
465	7.9	18	32	خواجة خليل /خام
452	7.8	18	4.8	المنتج
585	7.8	18	4.9	الحمداية /منتج
545	7.8	18	5	بعشيقه /منتج
463	7.96	18	22	الأيمن الموحد /الخام
454	7.8	18	4.1	المنتج
482	7.95	18	4.5	برطلة /منتج
486	7.8	18	55	الدندان /خام
476	7.7	18	4.9	المنتج
484	7.94	18	47	الاييسر القديم /خام
485	7.9	18	5	المنتج
480	7.93	18	4.83	التوسيع
472	7.8	18	4.94	بيسان
482	7.9	18	5	الحدياء
462	7.85	18	12	النمرود /خام
458	7.8	18	4.6	المنتج
508	7.9	18	50	الساهاون /خام
500	7.8	18	5	المنتج
455	7.6	18	13	الأيمن الجديد /خام
452	7.5	18	4.2	المنتج
567	8.1	18	4.88	بعشيقه
557	8.1	18	280	السلامية /خام
554	8	18	5	المنتج
498	7.94	18	250	الزهور /خام
485	7.85	18	5	المنتج
460	8	18	13.8	الاييسر الجديد /خام
457	7.9	18	4.75	المنتج
454	7.94	18	5	القبة
490	8	18	22	الدندان /خام
482	7.84	18	4.7	المنتج
508	7.9	18	4.8	الحمداية /منتج
487	7.9	18	4.8	الزهور /خام
482	7.73	18	3.5	المنتج
497	7.9	18	53	الساهاون /خام
491	7.8	18	5	المنتج

448	7.8	18	37	خواجة خليل /خام
443	7.7	18	4.5	المنتج
442	7.8	18	32	الأيمن الموحد /الخام
438	7.7	18	4	المنتج

البيانات المستخدمة بالبحث (شهر ٢/٢٠١٩)

E.C	PH	Tem	Tur	المشروع
497	8.1	19	4.8	الأيمن الموحد /الخام
426	7.9	19	30	المنتج
421	7.8	19	4.3	خواجة خليل /خام
504	8.25	19	4.9	المنتج
477	7.8	19	58	الحمداية /منتج
471	7.73	19	4	بعشيقه /منتج
440	7.8	19	26	الأيمن الموحد /الخام
437	7.6	19	2.9	المنتج
475	7.9	19	43	برطلة /منتج
468	7.8	19	3.2	الدندان /خام
428	7.7	19	22	المنتج
420	7.15	19	2.5	الايسر القديم /خام
510	8.3	19	5	المنتج
422	7.7	19	25	التوسيع
418	7.65	19	3	بيسان
468	8	19	39	الحدياء
464	7.85	19	3.2	النمرود /خام
458	8.2	19	27	المنتج
450	8.1	19	4.9	الساهاون /خام
432	7.7	19	20	المنتج
427	7.6	19	2.6	الأيمن الجديد /خام
466	7.9	19	23	المنتج
459	7.75	19	4.8	بعشيقه
514	8.2	19	4.9	السلامية /خام
432	7.9	19	18	المنتج
428	7.83	19	4.1	الزهور /خام
554	8	19	4	المنتج
457	7.8	19	18	الايسر الجديد /خام
450	7.7	19	3.9	المنتج
486	7.9	19	24	القبة
478	7.85	19	4.7	الدندان /خام
480	8	19	22	المنتج
476	7.85	19	4.9	الحمداية /منتج
428	7.9	19	20	الزهور /خام
418	7.8	19	4.6	المنتج

409	7.76	19	4.3	الساهاون /خام
427	7.8	19	4.7	المنتج
419	7.83	19	4.7	خواجة خليل /خام
469	8.1	19	4.8	المنتج
421	7.7	19	13	الأيمن الموحد /الخام
410	7.6	19	1.9	المنتج

البيانات المستخدمة بالبحث (شهر ٢٠١٩/٦)

E.C	PH	Tem	Tur	اسم المشروع
384	7.45	22	9	الايسر القديم /خام
380	7.33	22	4.1	المنتج
381	7.3	22	4.3	التوسيع /منتج
382	7.4	22	4.6	بيسان /منتج
382	7.4	22	4.6	الحدباء /منتج
540	7.15	22	9.6	الدندان خام + ابار
532	7.1	22	4.25	الدفع
380	7.2	22	5.8	الأيمن الموحد /خام
376	7.1	22	2.4	الدفع
478	7.3	22	20	الكوير /خام
470	7.1	22	5	المنتج
406	7.5	22	12	الساهاون /خام
397	7.35	22	3.8	المنتج
437	7.2	22	4.3	بعشيقه /منتج
402	7.35	22	8.5	الايسر القديم /خام
400	7.2	22	3.6	المنتج
397	7.25	22	4	التوسيع /منتج
398	7.3	22	4.3	بيسان /منتج
400	7.3	22	4.3	الحدباء /منتج
433	7.1	22	4.6	الحمدانية /منتج
988	7.3	22	4.8	برطلة ابار + منتج
483	11.2	22	11.2	الدندان خام + ابار
477	6.93	22	4.4	المنتج
394	7.1	22	5.2	الأيمن الجديد /خام
389	7	22	2.3	المنتج
416	7.4	22	11	الساهاون /خام
411	7.25	22	4.2	المنتج
349	6.9	22	4.25	المنتج
389	7.25	22	4	الحدباء /منتج
385	7.2	22	7.8	الأيمن الموحد /خام
381	7.1	22	3.2	المنتج

390	7.4	22	6.5	خواجة خليل /خام
384	7.3	22	3.8	المنتج
409	7.5	22	8	الزهور /خام
400	7.35	22	3.3	المنتج
460	7.6	22	12	الذندان /خام
456	7.44	22	4.25	المنتج
430	7.1	22	3.6	الحمدانية /منتج
413	7.5	22	10.8	الساهاون /خام
407	7.25	22	3.4	المنتج
399	7.1	22	6.8	الأيمن الجديد /خام
393	7	22	2.5	المنتج
790	7.3	22	4.7	بعشيفة + ابار
E.C	PH	Tem	Tur	اسم المشروع
435	7	22	3.8	الحمدانية /منتج
403	7.3	22	6.5	الرشيدية /خام
399	7.2	22	3.2	المنتج
992	7	22	4.2	برطلة /منتج + ابار
408	7	22	11.3	النمرود /خام
400	6.9	22	4.8	المنتج
398	7.3	22	8	الايبر القديم /خام
393	7.1	22	3.5	المنتج
495	7.25	22	3.7	التوسيع
396	7.18	22	3.9	بيسان /منتج
396	7.25	22	3.9	الحدباء /منتج
405	7.15	22	8.6	المنتج

البيانات المستخدمة بالبحث (شهر ٢٠١٩/٧)

E.C	PH	Tem	Tur	اسم المشروع
389	7.5	22	5.9	خواجة خليل
382	7.35	22	2.7	المنتج
399	7.1	22	5.5	الأيمن جديد/خام
390	7	22	2.2	المنتج
311	7.4	22	6	الزهور /خام
397	7.3	22	1.9	المنتج
411	7.5	22	4.6	المنتج
400	7.15	22	5.7	الأيمن الموحد/خام
389	7.1	22	2.3	المنتج
659	7	22	3.9	بعشيقية /المنتج + ابار
513	7.3	22	4.2	برطلة
413	7.5	22	8	الساهاون /خام
410	7.25	22	1.5	المنتج
470	7.4	22	10	الدندان /خام
464	7.2	22	3.5	المنتج
431	7.4	22	2.3	الحمدانية /المنتج
399	7.9	22	5.7	الأيمن الجديد /خام
391	7.8	22	1.8	المنتج
378	7.6	22	12	النمرود /خام
370	7.5	22	4.7	المنتج
405	7.8	22	9	القيارة /خام
402	7.7	22	4.2	المنتج
389	7.5	22	7.3	خواجة /خليل /خام
378	7.35	22	3.5	المنتج
795	7.11	22	0.13	محطة البعاج / ابار
401	7.8	22	5.8	الأيمن الموحد /خام
393	7.75	22	2.4	المنتج
395	7.4	22	5.6	الأيمن الجديد /خام
384	4.3	22	1.9	المنتج
384	7.7	22	8.7	الدندان /خام
471	7.6	22	4.1	المنتج
397	7.6	22	7.6	الايبر القديم /خام
394	7.45	22	3.5	قديم /منتج
396	7.4	22	4	التوسيع /منتج
393	7.5	22	3.9	الحديباء /المنتج
395	7.5	22	4.2	بيسان /المنتج
738	7.8	22	4.5	برطلة /منتج
393	7.5	22	6.8	الايبر الجديد /خام
389	7.3	22	2.9	المنتج



391	7.3	22	3.2	القبة /منتج
434	8	22	4.4	الحمدانية /المنتج
393	7.15	22	3.2	المنتج
402	7.8	22	5.2	الأيمن الجديد /خام
400	7.7	22	5.8	الايسر القديم /خام