تصنيف البيانات وقياس كفاءتها باستخدام أسلوب K-Means ونموذج VRSمديريات التربية العراقية - أنموذجا

م. م. احمد هشام محمد البصري
جامعة البصرة - كلية الإدارة والاقتصاد - قسم الإحصاء

Data classification and efficiency measurement using the K-Means method and the VRS model Directorate of Iraqi Education - a model

Ahmed Husham Mohammed Albasri
Assistant Lecturer, Department of statistics
College of Administration and Economics
University of Basrah

تصنيف البيانات وقياس كفاءتها باستخدام أسلوب K-Means ونموذج VRS مديريات التربية العراقية - أنموذجا

م. م.احمد هشام محمد البصري

الملخص :

يعد أسلوب K-Means من الأساليب الإحصائية المعتمدة في تصنيف البيانات من خلال تجميع البيانات على شكل عناقيد بحيث كل عنقود يمثل كياناً مستقلاً يحتوي على عدد من العوامل المتجانسة فيما بينها، ولكن آلية التصنيف لا يمكن أن تحدد وبشكل صريح نوع العنقود الذي يحتوي على المجاميع المصنفة، لذا هدفت الدراسة إلى إجراء تصنيف للعناقيد من خلال قياس كفاءتها بالنسبة لوحدات صنع القرار باستخدام نموذج البرمجة الخطية متعدد المعايير VRS (Variables Return Scales)، إذ طبقت هذه الدراسة على مديريات التربية في وزارة التربية العراقية–التعليم الابتدائي للعام الدراسي 2012–2013. إذ تمكنت الدراسة من خلال النتائج التوصل إلى مجموعتين (عنقودين) الأول تضمن(6) مديريات أما الثاني فقد تضمن (14) مديرية حسب أسلوب K-Means، كما تم تطبيق نموذج VRS لتحديد أكثر العناقيد كفاءة من خلال قياس الكفاءة الكلية لكلا العنقودين، إذ تبين أن العنقود الثاني هو الأفضل من خلال قياس كفاءة العنقود الكلية وباتجاهين مدخلات ومخرجات وبنسبة كفاءة 0.823 و 0.958 .

الكلمات الافتتاحية :تصنيف البيانات، أسلوب المتوسطات K-Means، البرمجة الخطية، أنموذج VRS

Abstract

The K-Means method is one of the statistical methods used to classify data by collecting data in the form of clusters so that each cluster represents an independent entity that contains a number of homogeneous factors among them. However, the classification mechanism can not explicitly determine the type of cluster containing classified aggregates . The study aimed at the classification of clusters by measuring their efficiency for decision-making units using the Variable Return Scales (VRS) model. This study was applied to the directorates of education at the Iraqi Ministry of Education - Primary Education for the academic year 2012-2013. The results of the study were based on the results of the first two clusters, which included (6) directorates. The second included (14) districts according to the K-Means method. The VRS model was applied to determine the most efficient clusters by measuring the total efficiency of both clusters. It was found that the second cluster is the best by measuring the efficiency of the total cluster and in two directions inputs and outputs with an efficiency of 0.823 and 0.958.

Keyword: Classification Data, K-Means, LP, VRS Model

المقدمـــة

أن تصنيف البيانات يُعد من الأساليب المهمة التي تسهم بشكل كبير في هيكلة البيانات وتجميعها على وفق معايير وضعت لغرض بيان أهميتها ودورها في إعداد مقارنات تسهم في بناء منهجية واضحة حول المجاميع المُشكلة من خلال تحديد مكامن التشابه والاختلاف فيما بينها، إن ولأهمية هذا الموضوع دعا الباحثون إلى أعداد دراسات أخذت على عاتقها تطوير اساليب التصنيف منذُ عام (1960) وليومنا هذا، ان أهمية النظام التعليمي على مختلف مستوياته قد والى اهتماماً كبيراً من قبل جميع الأنظمة للدور الكبير الذي يؤديه في خدمة المجتمع من خلال أعداد المورد البشري المتعلم، اذ يُعد التعليم الابتدائي الذي يطلق عليه أيضاً بالتعليم الأساسي من أهم المراحل في سلم التعليم والذي يقاس من خلاله التطور في جميع المجالات الاقتصادية والاجتماعية والثقافية ...الخ لذا حاولنا من خلال هذا البحث تقديم أحد اساليب التحليل العنقودي ألا وهو أسلوب المتوسطات K-means الذي يعرف أيضاً بالتحليل العنقودي غير الهرمي في دراسة هذا القطاع المهم من خلال تصنيف مديريات التربية في العراق تنبعاً لعدد من المعايير، إن خوارزمية هذا الأسلوب تبنى من خلال حساب المسافة الاقليدية بين الحالات على وفق صفاتها ثم إيجاد تشكيل العناقيد على وفق طريقة الجوار الأقرب، إن كفاءة العنقود تعتمد على قيم الأوساط الأولية المحسوبة لعوامل (متغيرات) الدراسة، ولكن في حالات يمكن أن تكون هناك اختلافات طفيفة بين الأوساط وهنا لا يمكن أن نجزم بالضرورة أن هذا التجميع هو أفضل من الآخر لذا في هذه الدراسة عمدنا إلى استخدام نموذج VRS الذي يُعد احد نماذج البرمجة الخطية متعددة المعايير ليكون داعماً لعملية التصنيف والوصول إلى نتائج أدق في اختيار العنقود .

1-مشكلة الدراسة

إن تصنيف البيانات يُعد من الضروريات المهمة التي تسهم وبشكل كبير في وضع الخطط والسياسات للنهوض بالقطاع المبحوث من خلال تجميع تلك البيانات في كيانات متجانسة أو متجانسة إلى حد ما، ولكن في اغلب الأحيان تكون هذه المجاميع (العناقيد) المشكلة غير معروفة كفاءتها، لذا يمكن أن نضع مشكلة الدراسة في كيفية تحديد نوع العناقيد المصنفة من خلال قياس كفاءة الحالات المصنفة .

2-هدف الدراسة

تهدف الدراسة إلى إعداد منهج علمي من خلال الدمج بين الأساليب الإحصائية أحدهما يصنف البيانات على شكل عناقيد والآخر يقيس كفاءة الحالات المصنفة لتحديد نوع كل عنقود، إذ تم استخدام أسلوب K-Means في تصنيف (عنقدة) مديريات التربية في وزارة التربية العراقية ثم تحديد نوع كل عنقود من خلال قياس كفاءة المديريات على وفق العوامل المحددة باستخدام نموذج (VRS) .

3-أهمية الدراسة

إن أهمية الدراسة تكمن في كيفية تصنيف مديريات التربية التابعة لوزارة التربية العراقية للعام الدراسي 2012-2013 وتجميعها على شكل عناقيد متجانسة ثم تحديد نوع كل عنقود من خلال قياس كفاءة تلك المديريات

التي تم تجميعها مما يتيح ذلك أمام صناع القرار في وضع الخطط والسياسات التي تسهم في معالجة الإشكاليات والنهوض بمستوى العمل في تلك المؤسسات .

4-مجتمع الدراسة

تم اعتماد بيانات مديريات التربية في وزارة التربية العراقية المعلنة على موقع وزارة التربية للعام الدراسي 2012-2013م .

5-الدراسات السابقة

من خلال هذه الفقرة سنتطرق إلى بعض الدراسات التي استهدفت هذا النوع من الأساليب وحسب تسلسلها الزمني:

(Redmond & Heneghan,2007) عمد الباحثون إلى اقتراح خوارزمية K-means إذ تم استخدام طريقة الشجرة (Kd) في تقدير كثافة البيانات في مواقع مختلفة بعد ذلك تم استخدم طريقة خوارزمية Katsavounidis، التي تدمج معلومات كثافة البيانات تم اختبار الخوارزمية المقترحة على 36 مجموعة من البيانات التركيبية وتمت مقارنتها مع 25 عملية من طرائق التهيئة العشوائية استنتجت الدراسة إلى أن اختبار Kd لخوارزمية K-means يوفر طريقة سريعة وموثوقة للتهيئة من خلال مقارنة الأسلوب المقترح مع 3 تقنيات أخرى للتهيئة في حين أن الطريقة فشلت في بعض الأحيان. **(Po & etc,2009)** اقترح الباحثون منهجاً جديداً في عنقدة وحدات صنع القرار من خلال قياس كفاءتها باستخدام أسلوب CCR باعتماد مدخلات ومخرجات النموذج المقترح وتوصلت الدراسة الى بناء خوارزمية حصينة لعنقدة الوحدات على وفق مدخلات ومخرجات النموذج المقترح . **(O. J & etc,2010)** هدفت الدراسة إلى اعتماد أسلوب التجميع غير الهرمي K-means في تصنيف ومراقبة تقدم أداء الطلبة الأكاديمي في التعليم العالي النيجيري بالاعتماد على تحليل نتائج الطلبة على وفق العلامات المتحققة لمستوى أدائهم، إذ بينت الدراسة أن خوارزمية التجميع تعمل كمقياس جيد لمراقبة تقدم أداء الطلاب في المؤسسة العليا. كما يعزز اتخاذ القرار من قبل المخططين الأكاديميين لرصد الفصل الدراسي للمرشحين حسب الفصل الدراسي عن طريق تحسين النتائج الأكاديمية المستقبلية في الدورة الأكاديمية اللاحقة. **(رشيد ومهدي، 2011)** تضمنت الدراسة تحليل واقع التربية والتعليم العالي للمحافظات العراقية غير المنتظمة بإقليم باستخدام اساليب التحليل العنقودي الهرمية وغير الهرمية إذ تضمنت الدراسة (25) متغيراً حاولت من خلالها تحديد تجمع المحافظات على وفق العوامل الدراسية المقترحة **(Celebi & etc, 2012)** هدفت الدراسة إلى تقديم نظرة عامة على الأساليب المعتمدة في العنقدة K-mean مع التركيز على كفاءتها الحسابية، ثم إعداد مقارنة بين ثماني طرق لتعيين تعقيد زمني خطي شائع الاستخدام في مجموعة كبيرة ومختلفة من مجموعات البيانات باستخدام معايير الأداء المختلفة، إذ بينت الدراسة أن خوارزمية K-Means شديدة الحساسية للوضع الأولي لمراكز العنقدة. **(القرشي و عبد الله، 2013)** يهدف البحث إلى الربط بين أسلوبي التحليل العنقودي والعمليات التصادفية (سلاسل ماركوف) لتقدير المصفوفة الاحتمالية الانتقالية، إذ عمد الباحثين إلى استخدام أسلوب المحاكاة لتوليد البيانات وإجراء عملية العنقدة وتقدير المصفوفة الانتقالية. **(Morissette & Chartier, 2013)** هدفت الدراسة إلى تعريف المنهج العام والتطبيق الرياضي لخوارزمية تجميع البيانات باعتماد تقنية K-means، إذ تمت مناقشة

هذه التقنية من خلال ثلاث خوارزميات مختلفة هي (Forgy/Lloyd، MacQueen، Hartigan & Wong) واستنتج الباحثان من خلال هذا المقترح أهمية طريقة K-mean في تقسيم المجموعات في أي حقل من الحقول البحثية. **(حسين، 2014)** قدمت الدراسة تحليلاً وصفياً للواقع التعليمي الأساسي في محافظة اللاذقية للمدة من 2006–2011 ودراسة تطورها خلال الزمن لتحديد المتطلبات المستقبلية للمدارس واستيعاب الأعداد المتزايدة من الطلبة، ومن خلال التحليل تمكنت الدراسة من التوصل إلى بعض الاستنتاجات منها تطور أعداد المدارس والشعب والتلاميذ وأعداد الهيأة التعليمية في الدراسة الأساسية بالنسبة للزمن، عدم الحاجة إلى أنشاء مدارس للتعليم الأساسي حتى عام 2016. **(حجوز، 2014)** تضمنت الدراسة تقديم مقترح لتحسين خوارزميات التعنقد بأسلوبK-Means بهدف إيجاد طرائق يمكن من خلالها معرفة المعلومات المخبأة ضمن الكم الهائل من البيانات، وتم ذلك من خلال الربط بين تحليل البيانات واستخدام أسلوب الذكاء الصناعي لتحقيق الكفاءة في التنقيب عن البيانات، ركز البحث على طريقة اختيار مركز العنقود لتحسين أداء العنقدة، واستخدام مراكز العناقيد الأولية لتحديد مركز العنقود الأفضل. **(إبراهيم، 2016)** صنفت الدراسة المحافظات العراقية على وفق بعض المؤشرات الصحية لعام 2010 وتحديد أي المحافظات متقدمة على الأخرى من حيث تقديمها للخدمات الصحية وأيها متأخر في تقديم تلك الخدمات، اقترحت الدراسة استخدام أسلوب التحليل العنقودي K-Means والمتمثلة بطريقة الجوار الأقرب في تشكيل العناقيد وتمكنت الدراسة من تحديد المحافظة المتقدمة من حيث تقديمها للخدمات الصحية .

6–الجانب النظري

التحليل العنقودي

يُعد التحليل العنقودي من تقنيات تصنيف البيانات وتحليلها والتي يمكن تطبيقها على مجموعات البيانات متعددة المتغيرات للكشف عن البنية الموجودة فيها(**Morissette,** 2013,Page 15)، إذ يعتمد هذا التحليل على إيجاد نقاط التشابه والاختلاف بين المتغيرات باعتماد أسس معينة لتحديد نوع العلاقة بين العوامل المطلوب تصنيفها(مصطفى، 2007، ص28)، ينقسم التحليل العنقودي إلى قسمين هما (التحليل العنقودي الهرمي Hierarchical Cluster Analysis) والتحليل العنقودي غير الهرمي أو ما يسمى بطريقة المتوسطات K-Means)(رشيد و مهدي، 2011،ص195)، لذا يمكن تعريف التحليل العنقودي بأنه "أسلوب تصنيف إحصائي يهدف إلى اكتشاف فيما إذا كانت العناصر (الحالات) في مجموعة من البيانات ينتمون إلى مجموعات مختلفة عن طريق إجراء مقارنات كمية من خلال عدد من العوامل" (Jain, 2010, Page652).

طريقة المتوسطات K-Means

إن طريقة المتوسطات تُعد من طرائق التعنقد غير الهرمية التي تم تطويرها خلال ثلاثة العقود الماضية وهي واحدة من أكثر الطرائق استخداماً في مجموعة متنوعة من المجالات البحثية (Khan & Ahmad, 2004, Page1293)، وهي طريقة شائعة الاستخدام تهدف إلى تقسيم البيانات تلقائياً إلى (K) من المجاميع من خلال تحديد نقطة المتوسط للمجاميع المعنقدة، ويتم ذلك من خلال حساب المسافة الاقليدية بين العوامل (Nerurkar & etc,2018,Page772) .

خطوات تطبيق طريقة المتوسطاتK-Means

1- حساب المسافة الاقليدية Euclidian distance

يُعد مقياس المسافة الاقليدية من أكثر المقاييس استخداماً في تقدير المسافة بين العوامل، فإذا كان لدينا P_j ; $j = 1,2, \dots, c$ هي عدد من العوامل معرفه على وفق مجموعة من الصفات $X_{P_j} = \{x_1, x_2, \dots, x_n\}$ وعليه يمكن حساب المسافة الاقليدية E_{dist} بين أزواج العوامل لكل صفة على وفق الصيغة التالية(القرشي و عبد الله، 2013، ص115) :

$$\boldsymbol{E_{dist_i} = \sqrt{\sum_{j=1}^{n} (X_{P_j} - X_{P_{j+1}})^2}} \qquad \dots (\mathbf{1})$$

تستخدم الصيغة المذكورة آنفاً كان عدد المشاهدات لصفات العوامل متساوياً، أما في حالة كون عدد المشاهدات غير متساوٍ، تستخدم الصيغة التالية (مصطفى، 2007،ص29):

$$\boldsymbol{E_{dist_i} = \sqrt{\frac{1}{p}\sum_{j=1}^{n} (X_{P_j} - X_{P_{j+1}})^2}} \qquad \dots (\mathbf{2})$$

إذ أن **p** : تمثل عدد العناصر، حيث ان **(p<n)**

2- عند الحصول على المسافات بين الحالات المدروسة باعتماد الصيغة (1) ولتحديد العناقيد الأولية يتم اختيار اقل مسافة بالنسبة للمجاميع المُشكلة على وفق الصيغة :

$$\boldsymbol{d_i = \min(E_{dist_i}) \;\; ; \;\; i = 1, 2, \dots, n} \qquad \dots (\mathbf{3})$$

3-حساب نقطة مركز التجميع الجديدة ولكل عنقود، على وفق الصيغة التالية :

$$\boldsymbol{c_k = \frac{\sum_{i=1}^{n} d_i}{s_k}} \; ; \qquad \dots (\mathbf{4})$$

إذ ان s_k : تمثل عدد الأعضاء في كل عنقود .

4- يتم تكرار الخطوة (2) و (3)، ويتم التوقف في حالة الحصول على مجاميع متشابهة (Khan & etc, 2004, Page1294)

اختبار الاختلافات بين متوسطات العوامل للعناقيد

يستخدم المختبر الإحصائي F للاختبار الاختلافات بين متوسطات العوامل للعناقيد المشكلة من خلال الحصول على مركبات جدول تحليل التباين وهي مجموع المربعات الكلية(TSS) ومجموع المربعات المشروحة(ESS) ومجموع مربعات البواقي(RSS)، (One-Way ANOVA test) :

$$Tss = \sum_{c=1}^{k} \sum_{i=1}^{n} (X_{ci} - \bar{X}_{..})^2 \qquad \dots (\mathbf{5})$$

$$Ess_c = \sum_{i=1}^{n_c} (X_{ci} - \bar{X}_c)^2 \; ; \; c = 1,2, \dots, k \qquad \dots (\mathbf{6})$$

$$RSS = Tss - Ess_c \qquad \dots (\mathbf{7})$$

(**Ostertagová** & etc, 2013, Page 257)

7-نموذج VRS (Variable Return Scale)

ويسمى بنموذج عائد الحجم المتغير وهو من نماذج البرمجة الخطية متعدد المعايير باتجاه x_{ij} و y_{ij} التي تمثل مدخلات ومخرجات النموذج، الذي يستخدم في قياس الكفاءة في ظل تحديد عائد الحجم (متناقص، ثابت، متزايد)، إذ طور نموذج CRS عام (1984) من خلال إدخال بعض التعديلات في قيوده لمعالجة عوائد الحجم المتغيرة (Ramanathan, 2003, Page 69)، تم تعديل النموذج بإضافة قيد التحدب $\sum_{j=1}^{n}\lambda_j = 1\ ;\ \lambda_j \geq 0$ (Cooper & etc, 2007, Page 91)، ان التحليل الحدودي لنموذج VRS يتم باتجاهين هما (النموذج الادخاليIVRS، والنموذج الاخراجيOVRS) :

- النموذج الادخاليIVRS (Input Variable Return Scale)

$$Min\ \theta$$
Subject to:
$$\sum_{j=1}^{n}\lambda_j x_{ij} \leq \theta x_o$$
$$\sum_{j=1}^{n}\lambda_j\, y_{rj} \geq y_{ro} \qquad \dots(\mathbf{8})$$
$$\sum_{j=1}^{n}\lambda_j = 1$$
$$\boldsymbol{\lambda_j \geq 0}$$

- والنموذج الاخراجيOVRS (Output Variable Return Scale)

$$Max\ \phi$$
Subject to:
$$\sum_{j=1}^{n}\lambda_j\, y_{rj} \geq \phi y_{ro}$$
$$\sum_{j=1}^{n}\lambda_j\, x_{ij} \leq x_{io} \qquad \dots(\mathbf{9})$$
$$\sum_{j=1}^{n}\lambda_j = 1$$
$$\boldsymbol{\lambda_j \geq 0}$$
(Widiarto & etc, 2017, Page 281)

إذ أن :

x_{ij} : تمثل مدخلات النموذج للعامل i ولكل وحدة حدة صنع قرار j .

y_{ij} : تمثل مخرجات النموذج للعامل i ولكل وحدة حدة صنع قرار j .

x_{ro} : يمثل الطاقة المتاحة من مدخل الوحدة المستهدفة .

y_{ro} : يمثل الطاقة المتاحة من مخرج الوحدة المستهدفة .

وعليه يمكن أن نحدد الافتراضات التي يتوقف عليها تقدير الكفاءة لنموذج VRS :

1- $1 \leq \sum_{j=1}^{n} \lambda_j$; $\forall \lambda_j \geq 0$، إذا كان مقياس الكفاءة يقع تحت هذا الافتراض فانه يمثل تناقص عائد الحجم .

2- $\sum_{j=1}^{n} \lambda_j = 1$; $\forall \lambda_j \geq 0$، إذا كان مقياس الكفاءة يقع تحت هذا الافتراض فانه يمثل ثبات عائد الحجم

3- $\sum_{j=1}^{n} \lambda_j \geq 1$; $\forall \lambda_j \geq 0$، إذا كان مقياس الكفاءة يقع تحت هذا الافتراض فانه يمثل تزايد عائد الحجم (Kutin & etc, 2017, Page 70)

8-الجانب العملي

تم تطبيق الدراسة على بيانات مديريات التربية التابعة لوزارة التربية العراقية للعام الدراسي 2012 – 2013 وذلك من خلال البيانات المنشورة على موقع الوزارة وباستعمال برنامج XLDEA. وتم تجميع البيانات الخاصة بالدراسة الابتدائية بمواقع (20) مديرية تربية، أما متغيرات الدراسة فتمثلت (عدد المدارس، عدد الطلبة، عدد الهيأت التعليمية والتدريسية، عدد الشعب، عدد الأبنية، نسب النجاح، نسب التسرب) .

جدول (1)

بيانات التعليم الابتدائي لمديريات التربية العراقية غير المنتظمة بإقليم

ID	المحافظة	القائمقامية	عدد المدارس	عدد الطلبة	عدد المعلمين والمعلمات والتدريسيين	عدد الشعب	عدد الأبنية المدرسية	نسبة النجاح	نسب التسرب
1	نينوى	نينوى	1679	548812	19740	14944	1369	88.9	0.025
2	صلاح الدين	صلاح الدين	1145	277206	14992	9536	849	96.2	0.016
3	كركوك	كركوك	1051	205212	12225	8886	803	92.8	0.013
4	ديالى	ديالى	884	258791	16910	8136	729	86.3	0.006
5	بغداد	الرصافة 1	354	195842	11426	4873	221	85.6	0.015
6	بغداد	الرصافة 2	472	360117	12964	7293	327	93.8	0.02
7	بغداد	الرصافة 3	331	183696	7835	4841	210	96.3	0.025
8	بغداد	كرخ الأولى	264	115657	9021	3378	233	80.6	0.018
9	بغداد	كرخ الثانية	492	234395	13180	6462	356	97.3	0.019
10	بغداد	كرخ الثالثة	270	171003	8907	4397	245	93.5	0.016
11	الانبار	الانبار	1162	329439	18139	11284	797	96.4	0.02
12	بابل	بابل	853	354545	17775	9397	600	96.8	0.024
13	كربلاء	كربلاء	485	212169	11966	5722	275	84.7	0.018
14	النجف	النجف	566	245105	11784	6804	374	98.7	0.018
15	القادسية	القادسية	671	218838	13427	6512	512	90.4	0.018
16	المثنى	المثنى	475	134661	7422	4372	378	94.6	0.013
17	واسط	واسط	799	219066	12796	7186	616	95.1	0.021
18	ذي قار	ذي قار	1203	354079	20314	10478	893	91	0.014
19	ميسان	ميسان	631	187945	12763	6272	486	92.5	0.014
20	البصرة	البصرة	1043	482267	19709	11526	600	83.3	0.02

المصدر : بيانات وزارة التربية العراقية للعام الدراسي 2012-2013 – منشورة على الموقع الالكتروني الرسمي

http://www.moedu.gov.iq/upload/upfile/ar/131.xlsx

مناقشة نتائج K-Means

يوضح الجدول (2) المجاميع المعنقدة، والمسافات المقدرة بين كل تربية حسب العنقود التي تم تقديرها حسب الصيغة (1)، إذ تبين النتائج انه قد تم تشكيل عنقودين يحتوي كل عنقود على مجموعة متجانسة من المديريات تبعاً لما متحقق لديها من مدخلات ومخرجات، ولتوضيح نتائج كل عنقود ستتم تجزئة المجاميع حسب كل عنقود .

جدول (2)
يوضح المجاميع المعنقدة

Case Number	التربية	Cluster	Distance
1	نينوى	1	144006.382
2	صلاح الدين	2	73098.784
3	كركوك	2	2907.122
4	ديالى	2	54811.599
5	الرصافة الأولى	2	8537.871
6	الرصافة الثانية	1	45197.908
7	الرصافة الثالثة	2	20981.510
8	الكرخ الأولى	2	88688.637
9	الكرخ الثانية	2	30173.393
10	الكرخ الثالثة	2	33428.571
11	الانبار	1	75438.997
12	بابل	1	50353.445
13	كربلاء	2	7935.288
14	النجف	2	40852.825
15	القادسية	2	14679.518
16	المثنى	2	69755.471
17	واسط	2	14878.262
18	ذي قار	1	50846.921
19	ميسان	2	16341.988
20	البصرة	1	77410.478

المصدر: من أعداد الباحث حسب نتائج الدراسة باستعمال برنامج SPSS V.23

يتضح من الجدول (3) أن العنقود الأول قد تضمن (6) مديريات هي (نينوى، البصرة، الانبار، ذي قار، بابل، الرصافة الثانية)، رتبت حسب بعدها عن مركز العنقود حيث أن مديرية تربية نينوى تمثل المسافة الأبعد عن مركز العنقود والتي قدرت بـ(144006.382) أما مديرية تربية الرصافة الثانية فهي الأقرب من مركز العنقود والتي قدرت بـ(45197.908)، .

جدول (3)
مديريات التربية للعنقود الأول

Case Number	التربية	Cluster	Distance
1	نينوى	1	144006.382
2	البصرة		77410.478
3	الانبار		75438.997
4	ذي قار		50846.921
5	بابل		50353.445
6	الرصافة الثانية		45197.908

المصدر: من أعداد الباحث حسب نتائج الدراسة

يتضح من الجدول (4) أن العنقود الثاني قد تضمن (14) مديرية (الكرخ الأولى، صلاح الدين، ...، كركوك)، رتبت حسب بعدها عن مركز العنقود حيث أن مديرية تربية الكرخ الأولى تمثل المسافة الأبعد عن مركز العنقود والتي قدرت بـ(88688.637)، أما مديرية تربية كركوك فهي الأقرب من مركز العنقود والتي قدرت بـ(2907.122) .

جدول (4)
مديريات التربية للعنقود الثاني

Case Number	التربية	Cluster	Distance
1	كرخ الأولى	2	88688.637
2	صلاح الدين		73098.784
3	المثنى		69755.471
4	ديالى		54811.599
5	النجف		40852.825
6	الكرخ الثالثة		33428.571
7	الكرخ الثانية		30173.393
8	الرصافة 3		20981.510
9	ميسان		16341.988
10	واسط		14878.262
11	القادسية		14679.518
12	الرصافة الأولى		8537.871
13	كربلاء		7935.288
14	كركوك		2907.122

المصدر: من أعداد الباحث حسب نتائج الدراسة

يتضح من نتائج جدول (5) أن هناك اختلافات بين الأوساط الأولية للعوامل المبحوثة للعنقود الأول والثاني وبغية التأكد، من ذلك سيتم استخدام المختبر الإحصائي F كما موضح بالجدول (6) :

جدول (5)
الأوساط الأولية للعناقيد

Initial Cluster Centers		
Cluster		
2	1	
264	1679	عدد المدارس
115657	548812	عدد الطلبة
9021	19740	عدد الهيئات التعليمية والتدريسية
3378	14944	عدد الشعب
233	1369	عدد الأبنية المدرسية
80.60	88.90	نسبة النجاح
0.02	0.03	نسب التسرب

المصدر: من أعداد الباحث حسب نتائج برنامج SPSS V.23

يوضح جدول (6) تحليل التباين بين متوسطات العوامل للعناقيد المشكلة، إذ تم تقدير مركبات الجدول على وفق الصيغ (5، 6، 7)، إذ تبين قيم (F) لكل من متغير (عدد المدارس، عدد الأبنية المدرسية، نسب التسرب، ونسب النجاح) وهي (9.047، 5.905، 3.833، 0.000) على التوالي كانت قليلة وهذا يدل على أن الاختلافات بين العنقودين قليلة جداً، أما بالنسبة لقيم (F) لكل من (عدد الطلبة، عدد الهيأت التدريسية، عدد الشعب) كانت كبيرة بما يدل على أن هناك اختلافات كبيرة بين العنقودين، مع وجود هذه الاختلافات فلا يمكن أن نحدد أفضلية كل عنقود عن الأخر لذا تم استخدام نموذج VRS .

جدول (6)
تحليل الاختلافات بين أوساط عوامل العناقيد

ANOVA						
	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
عدد المدارس	917468.810	1	101413.899	18	9.047	.008
عدد الطلبة	169043816340.536	1	3635634917.956	18	46.496	.000
عدد الهيأت التعليمية والتدريسية	169132322.917	1	7263607.491	18	23.285	.000
عدد الشعب	88066991.260	1	4126171.761	18	21.344	.000
عدد الأبنية المدرسية	417438.288	1	70690.348	18	5.905	.026
نسب النجاح	.014	1	28.803	18	.000	.983
نسب التسرب	.000	1	.000	18	3.833	.066

المصدر: من أعداد الباحث نتائج الدراسة باستعمال برنامج SPSS V.23

مناقشة نتائج نموذج VRS

لتحديد كفاءة العنقود سيتم قياس كفاءة مديريات التربية للتعليم الابتدائي، بتطبيق الصيغة (8) و (9) .

يتضح من الجدول (7) أن الكفاءة الكلية للعنقود الأول حسب نموذج IVRS قد حقق مانسبته (0.57) وهي نسبة ضعيفة باتجاه مدخلات العملية التعليمية هذا يعني انه يتطلب إلى تحسين أداء اغلب مديريات التربية

في هذا العنقود وهي (ذي قار، نينوى، البصرة، الانبار، الرصافة الثانية) من خلال الاستثمار الأمثل لمواردها لتحقيق الكفاءة التامة أما مديرية تربية بابل فقد حققت الكفاءة التامة، بينما الكفاءة الكلية للعنقود الأول حسب نموذج OVRS قد حققت مانسبته (0.95) باتجاه مخرجات العملية التعليمية أي أنها تحتاج إلى تحسين طفيف لكي تصل إلى الكفاءة التامة.

جدول (7)
يوضح الكفاءة الكلية للعنقود الأول للتعليم الابتدائي

Case Number	التربية	IVRS	OVRS
1	نينوى	0.3969	1
2	البصرة	0.4002	0.8586
3	الانبار	0.5607	0.984
4	ذي قار	0.3929	0.922
5	بابل	1	1
6	الرصافة الثانية	0.6654	0.9598
الكفاءة الكلية للعنقود		0.57	0.95

المصدر: من أعداد الباحث حسب نتائج الدراسة باستعمال برنامج XLDEA

يتضح من الجدول (8) أن الكفاءة الكلية للعنقود الثاني حسب نموذج IVRS قد حققت مانسبته (0.823) وهي نسبة جيدة باتجاه مدخلات العملية التعليمية ولكن للوصول للكفاءة التامة يتطلب تحسين أداء بعض مديريات التربية المحددة في العنقود وهي مرتبة تصاعدياً من الأضعف إلى الأقوى هي (ديالى، صلاح الدين، كركوك، القادسية، ميسان، واسط، كربلاء، الكرخ الثانية، الرصافة الأولى) من خلال الاستثمار الأمثل لمواردها لتحقيق الكفاءة التامة، بينما الكفاءة الكلية للعنقود الثاني حسب نموذج OVRS قد حققت مانسبته (0.958) باتجاه مخرجات العملية التعليمية أي أنها تحتاج إلى تحسين طفيف في مخرجاتها لكي تصل إلى الكفاءة التامة، أما بالنسبة إلى مديريات تربية (الكرخ الأولى، المثنى، النجف، الكرخ الثالثة، الرصافة الثالثة) قد حققت الكفاءة التامة في كلا النموذجين .

جدول (8)
يوضح الكفاءة الكلية للعنقود الثاني للتعليم الابتدائي

Case Number	التربية	IVRS	OVRS
1	الكرخ الأولى	1	1
2	صلاح الدين	0.642	0.975
3	المثنى	1	1
4	ديالى	0.485	0.874
5	النجف	1	1
6	الكرخ الثالثة	1	1
7	الكرخ الثانية	0.892	0.993
8	الرصافة 3	1	1
9	ميسان	0.714	0.958
10	واسط	0.756	0.975
11	القادسية	0.649	0.925
12	الرصافة الأولى	0.961	0.889
13	كربلاء	0.785	0.871
14	كركوك	0.644	0.955

الكفاءة الكلية للعنقود	0.823	0.958

المصدر: من أعداد الباحث حسب نتائج الدراسة باستعمال برنامج XLDEA

الاستنتاجات والتوصيات

لاستنتاجات

من خلال النتائج المتحققة في الجانب التطبيقي يمكن أن نبين عدداً من الاستنتاجات وهي :

1- أهمية تطبيق أسلوب K-Mean في تصنيف البيانات وتشكيلها بما يضمن تحقيق الفائدة للوصل إلى أفضل تجميع للبيانات .

2- أهمية استخدام نموذج VRS في تحديد نوع التصنيف للعناقيد المشكلة في حالة وجود اختلافات غير واضحة بين متوسطات العوامل المبحوثة .

3- تمكنت الدراسة من قياس كفاءة مديريات التربية العراقية للمحافظات غير المنتظمة بإقليم .

4- أوضحت نتائج الدراسة أن العنقود الثاني هو الأفضل من خلال قياس الكفاءة الكلية للعنقود باعتماد الكفاءة المتحققة لكل مديرية .

التوصيات

1- توصي الدراسة باعتماد أساليب إحصائية حديثة تسهم في بناء خوارزمية تنظيمية لتهيئة البيانات وكيفية الاستفادة منها مثل أسلوب تنقيب البيانات Data Mining .

2- توصي الدراسة بأهمية متابعة عمل هذه المديريات بالنسبة لأهمية الخدمة التي تقدمها للمجتمع .

3- توصي الدراسة إلى تحسين أداء بعض المديريات باتجاه الاستثمار الأمثل لمواردها وتحقيق أفضل المخرجات .

المصـادر:

المصادر العربية :

1- إبراهيم، عمر سالم، (2016)، "استخدام المؤشرات الصحية لعام 2010 لتصنيف محافظات العراق باستخدام التحليل العنقودي"، مجلة تكريت للعلوم الصرفة، المجلد(21)، العدد (4)، ص(149– 159) .

2- القرشي، إحسان كاظم وعبد الله، ماجد جاسب، 2013، "التقدير اللامعلمي لمصفوفة الاحتمالات الانتقالية باستخدام التحليل العنقودي"، مجلة ميسان للدراسات الأكاديمية المجلد(12)، العدد(22)، ص(114– 122) .

3- حجوز، محمد مصطفى، 2014، "تحسين خوارزميات K-Means"، مجلة جامعة البعث، المجلد (36)، العدد(6)، ص(129 – 153) .

4- حسين، محمود حسن، 2014، "دراسة تحليله لواقع مدارس التعليم الأساسي وتطورها في محافظة اللاذقية خلال المدة (2006 – 2011)، مجلة جامعة دمشق للعلوم الاقتصادية والقانونية، المجلد(30)، العدد (1)، ص(585 – 646) .

5- رشيد، أسيل عبد الرزاق و مهدي، نبأ نعيم، 2011، "تحليل واقع التربية والتعليم في العراق باستخدام طرائق التحليل العنقودي"، مجلة القادسية للعلوم الإدارية والاقتصادية، المجلد(13)، العدد(2)، ص(194 – 217) .

6- مصطفى، نزار، 2007، "استخدام بعض طرق التحليل العنقودي في التصنيف مع تطبيق عملي"، هيئة التعليم التقني، مجلة التقني، المجلد(20)، العدد(2)، ص (27 – 37) .

المصادر الأجنبية:

7- Celebi, M. Emre; Kingravi, Hassan A.; Vela, Patricio A., 2013, "A Comparative Study of Efficient Initialization Methods for the K-Means Clustering Algorithm", Expert Systems with Applications, Vol.(40), No.(1), Pages(200 – 210) .

8- Cooper, William W.; Seiford, Lawrence M.; Tone, Kaoru, 2007, "Data Envelopment Analysis - A Comprehensive Text with Models, Applications, References and DEA-Solver Software", 2Ed, Springer Science Business Media, LLC, Pages(1 – 483) .

9- Jain, Anil K., 2010," Data clustering: 50 years beyond K-means", Pattern Recognition Letters, ScienceDirect, Vol. 31, No. 8, Pages(651-666) .

10- Khan, Shehroz S.; Ahmad, Amir, 2004," Cluster center initialization algorithm for K-means clustering", Pattern Recognition Letters, ScienceDirect, Vol. (25), No.(11), Pages(1293-1302)

11- Kutin, Nikola; Nguyen, Thanh Thuy; Vallee, Thomas, 2017,"Relative Efficiencies of ASEAN container ports based on data Envelopment Analysis",

The Asian Journal of Shipping and Logistic, Elsevier, ScienceDirect, Vol.(33), No.(2), Pages (67 – 77) .

12- Morissette, Laurence; Chartier, Sylvain, 2013," The k-means clustering technique: General considerations and implementation in Mathematica", Tutorials in Quantitative Methods for Psychology, Vol.(9), No.(1), Pages (15 – 24) .

13- Nerurkar, Pranav; Shirke, Archana; Chandane, Madhav; Bhirud, Sunil, 2018,"Empirical Analysis of Data Clustering Algorithms", Procedia Computer Science, ScienceDirect, Vol.(125), Pages(770 – 779) .

14- O. J, Oyelade; O. O, ladipupo; I. C, Obagbuwa, 2010," Application of k-Means Clustering algorithm for prediction of Students' Academic Performance", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, _o. 1, Pages (292 – 295) .

15- Stertagova, Eva; Ostertag, Oskar, 2013," Methodology and Application of One-way ANOVA", American Journal of Mechanical Engineering, Vol. (1), No.(7), Pages(256-261) .

16- Ramanathan, R, 2003," An Introduction to Data Envelopment Analysis A Tool for Performance Measurement", 1Ed, Published by Tejeshwar Singh for Sage Publications India Library of Congress Cataloging-in-Publication Data, Pages(1 – 180) .

*17- Redmond, Stephen J.; Heneghan, Conor, 2007," A method for initialising the K-means clustering algorithm using kd-trees", **Pattern Recognition Letters**, Elsevier, **Vol.(28), No.(8)**, Pages(965-973) .*

18- Po, Rung-Wei; Guh, Yuh-Yuan; Yang, Miin-Shen, 2009," A new clustering approach using data envelopment analysis", European Journal of Operational Research, Elsevier, Vol.(199), No.(1), Pages(276 – 284) .

19- Widiarto, Indra; Emrouznejad, Ali, Anastasakis, Leonidas, 2017,"Observing choice of loan methods in not-for-profit microfinance using data envelopment analysis", Expert Systems With Applications, Elsevier, Vol.(82), Pages(278 – 290) .

الموقع الالكتروني

20- **بيانات وزارة التربية العراقية للعام الدراسي 2012-2013 – منشورة على الموقع الالكتروني الرسمي**
http://www.moedu.gov.iq/upload/upfile/ar/131.xlsx