

## أنموذج تصادفي مضرب لعملية ماركوفية

## A Fuzzy Stochastic Model for a Markov Process

أ.م.د. مثنى صبحي سليمان م.م. شيماء وليد محمود م.م. غالية توفيق بشير

كلية علوم الحاسوب والرياضيات/ جامعة الموصل

## المستخلص

تم في هذا البحث الجمع بين نظرية متسلسلة ماركوف والتحليل العنقودي الذي يحدد مدخلات النموذج الماركوفي من الحالات، وتم الربط بين هذين النموذجين من خلال اقتراح نموذج ماركوفي مصاغ على أساس مبادئ العنقدة المضببة، والمقارنة مع نموذج ماركوفي مصاغ على أساس العنقدة غير المضببة. كما تم الاعتماد على خوارزمية K-Means للعنقدة وتم تضبيبيها لإجراء المقارنة. وكجانب عملي للبحث تم التطبيق على نسبة السرعات الحرارية في الفواكه والخضراوات، تم ملاحظة أن مصفوفة التوزيع المراوح للحالات للنموذج الماركوفي المصاغ على أساس العنقدة المضببة استقرت بشكل أسرع من النموذج الماركوفي المصاغ على أساس العنقدة غير المضببة، كما تم ملاحظة نسبة استقرار الفواكه والخضراوات قيد الدراسة على صفة انخفاض السرعات الحرارية اكبر من استقرارها على صفة ارتفاع السرعات الحرارية.

*الكلمات الدالة: متسلسلة ماركوف، العنقدة المضببة.*

## Abstract

In this paper, a combination of sequential Markov theory and cluster analysis, which determines inputs the Markov model of states, was the link between these two models by proposing Markov model formulated based on the principles of clustering Fuzzy, and comparison with the Markov model formulated based on clustering. It was also rely on the algorithm of K-Means clustering was Fuzzification to make a comparison. The practical side was applied to the caloric ratio in fruits and vegetables, it was noted that the Stationary distribution matrix of states of the Markov model formulated on the basis of clustering Fuzzy stabilized faster than the Markov model formulated on the basis of clustering, as was observed ratio stability of fruits and vegetables under study on low-calorie attribute is greater than the stability of the high-calorie attribute.

*Keywords: Markov chain, Fuzzy Clustering.*

## 1 المقدمة: Introduction

تعد النمذجة الماركوفية احد أنواع النمذجة الرياضية، التي تتعامل مع الظواهر المختلفة التي يكون حاضرها معتمداً على ما سبقه. ويتمتع النموذج الماركوفي بمكانة مرموقة في العلوم الرياضية والإحصائية، وقد ازدادت أهمية هذا النموذج بعد دخوله مجالات مهمة في علوم الحاسوب والمعلوماتية. [الخياط، 2011]

وتحتل نظرية ماركوف Markov Theory مكانة كبيرة وهامة جدا في العمليات التصادفية Stochastic Processes، ويعود ذلك إلى تعدد التطبيقات التي تتمتع بها هذه النظرية في النماذج الاقتصادية والزراعية والطبية وعلم الاجتماع وعلم الإدارة. لقد استطاع العالم Markov في عام 1907 من وضع المفاهيم الأساسية لمتسلسلة ماركوف والتي طورت من قبل العديد من الباحثين في تفسير سلسلة ماركوف على أنها عبارة عن متتابعة من الحالات التي يمكن أن يكون فيها نظام ما عند أي لحظة زمنية  $t$ ، أو متتابعة من المواضع التي يحتلها جسيم متحرك. ومن أهم أسباب الاهتمام الخاص بهذا النموذج كونه يتعامل مع ما يعرف بالحالات (states) التي من الممكن أن تكون أي صفة اسمية أو رتيبه أو رمز أو حرف أو أي شيء آخر. [الخياط، 2011]

إن هدف البحث هو الجمع بين نظرية متسلسلة ماركوف والتحليل العنقودي Cluster Analysis الذي يحدد مدخلات النموذج الماركوفي من الحالات (states)، ويتم الجمع أو الربط بين هذين النموذجين في هذا البحث من خلال اقتراح نموذج ماركوفي مصاغ على أساس مبادئ العقدة المضببة Fuzzy Clustering، والمقارنة مع نموذج ماركوفي مصاغ على أساس العقدة غير المضببة. إذ تم الاعتماد على خوارزمية K-Means للعقدة.

لقد قام الباحثان [Fathi and Gharehdaghi, 2012] باستخدام المتتالية العشوائية Halton Guasi لتوليد بعض متسلسلات ماركوف المضببة، ومقارنتها مع متتالية عشوائية أخرى تم توليدها بدالة عشوائية rand باستخدام اللغة البرمجية Matlab. كما اقترح الباحثان [Fathi and Masoumi, 2012] بإجراء محاكاة لمتسلسلة ماركوف باستخدام أحجام مختلفة ودوال عشوائية rand وأيضا باستخدام اللغة البرمجية Matlab. واستخدم الباحثان [Fathi and Gharehdaghi, 2014] متتالية Faure العشوائية ومتتالية Kronecker العشوائية لتوليد القيم العضوية لمتسلسلة ماركوف المضببة.

## 1 متسلسلة ماركوف: Markov chain

تعد متسلسلة ماركوف حالة خاصة من العملية التصادفية التي يكون فيها فضاء الحالة متقطع وتحمل الخاصية الماركوفية عندما يعتمد المستقبل على الحاضر فقط ولا يعتمد على الماضي وحسب الصيغة أدناه:

$$P(X_{t+1} = j | X_t = 0, \dots, X_0 = k) = P(X_{t+1} = j | X_t = i) = P_{ij} \quad \dots(1)$$

كما إن مجموعة القيم الممكنة للعملية التصادفية تسمى فضاء الحالة State Space، وإن فضاء الحالة يكون على نوعين محدوداً أو غير محدود حسب عدد عناصره. كذلك فإنه ممكن أن يكون مستمراً إذا اخذ قيماً متصلة، أو منقطعاً إذا اخذ قيماً منفصلة. من هنا يتضح بان هوية العملية التصادفية تتحدد من خلال فضاءين هما فضاء الحالة وفضاء المعلمة. [ الخياط، 2011 ]

إن العملية التصادفية بنيت على أساس انتقال الظاهرة من حالة إلى حالة أخرى استناداً إلى قوانين احتمالية معينة تدعى بالاحتمالات الانتقالية Transition Probability التي تمثل احتمال الانتقال خلال فترة زمنية معينة نرمز لها بالرمز  $P_{ij}$ ، ويمكن تمثيل التوزيع الاحتمالي للانتقالات بين الحالات بمصفوفة تسمى المصفوفة الانتقالية Transition Matrix التي هي عبارة عن مصفوفة تصادفية لها المميزات الآتية: [ احمد، 2008 ]

1. جميع عناصرها غير سالبة ( لكونها احتمالات).
2. مجموع كل صف من صفوفها يساوي الواحد (لان مجموع الاحتمالات الكلية يساوي الواحد).
3. يجب أن تكون مصفوفة  $P$  مربعة من الدرجة  $n \times n$  وكما موضحة أدناه:

$$P = \begin{matrix} s_1 \\ s_2 \\ \cdot \\ \cdot \\ s_n \end{matrix} \begin{bmatrix} p_{11} & p_{12} & \cdot & \cdot & p_{1n} \\ p_{21} & p_{22} & \cdot & \cdot & p_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{n1} & p_{n2} & \cdot & \cdot & p_{nn} \end{bmatrix}$$

### (3) تصنيف حالات سلاسل ماركوف: [Marius, 1980]

- (أ) سلسلة ماركوف غير قابلة للتجزئة Irreducible: هي السلسلة التي إذا أمكن الانتقال من أي حالة من حالاتها إلى الحالات الأخرى وبالعكس عند أي زمن.
- (ب) سلسلة ماركوف ذات العودة الموجبة Positive Recurrent: يقال للحالة  $i$  في سلسلة ماركوف بأنها حالة عودة موجبة إذا وفقط إذا كان من المؤكد رجوع العملية للحالة نفسها والتي سبق أن غادرتها، أي إذا كانت العودة النهائية إليها مؤكدة.
- (ج) الحالة الدورية Periodic: يقال للحالة  $i$  بأنها دورية إذا كان القاسم المشترك الأعظم لعدد الدورات التي تظهر فيها الحالة  $i$  اكبر من الواحد الصحيح وعكس ذلك تسمى الحالة  $i$  حالة غير دورية.
- (د) الحالة الثبوتية Ergodic: يقال للحالة  $i$  بأنها حالة ثبوتية إذا كانت غير قابلة للتجزئة وذات عودة موجبة وغير دورية.

هـ) التوزيع المتزن أو التوزيع المراح  $Stationary$  Distribution: هو التوزيع الاحتمالي لحالات سلسلة ماركوف بعد زمن طويل ويكون توزيع وحيد إذا كانت المصفوفة الانتقالية  $P$  تمثل سلسلة ثبوتية ويرمز له بالرمز  $\pi$  ويمكن إيجاد هذا التوزيع بحل النظام الآتي:

$$\pi P = \pi \quad \dots(2)$$

$$\sum_{i=1}^n \pi_i = 1$$

## Cluster Analysis

## (4) التحليل العنقودي:

بعد أن اكتشف الإنسان وجود صفات مشتركة في أجزاء المعرفة عني بتصنيفها، وبذل الفلاسفة والمفكرون جهوداً لوضع أسس ونظم لهذا التصنيف، وبهذا فإن كلمة "التصنيف" في اللغة هي تمييز الأشياء بعضها عن بعض بطريقة مرتبة في أصناف أو أقسام، وإذا عرفت ورتبت يكون قد تم إعداد نوعاً من أنظمة التصنيف. وعليه يمكن تعريف التصنيف بأنه عملية تجميع الأشياء المتشابهة معاً ويشترك جميع أعضاء المجموعة الواحدة أو القسم أو الصنف الواحد الناتجة عن التصنيف في خاصية محددة واحدة على الأقل لا يملكها أعضاء الأقسام أو الأصناف الأخرى للاستفادة منه في مجالات علمية. [يوسف، 2011]

من المعروف أن طرق التصنيف تقسم إلى مجموعتين رئيسيتين من التقنيات هما التحليل العنقودي Cluster Analysis والتحليل التمييزي Discriminant Analysis. إذ يعد التحليل العنقودي من التقنيات بدون إشراف، حيث أن المعلومات حول مجموعة التصنيف تكون قليلة أو معدومة، والهدف هو إيجاد مجموعات من البيانات. أما التحليل التمييزي فهو وجود علاقة عضوية في مجموعة عينات التدريب والغرض الرئيسي هو بناء قاعدة تصنيف ملائمة لعينات جديدة غير معروفة. [Naes and Movik, 1999]

ويعد التحليل العنقودي أحد طرائق تحليل المتغيرات المتعددة، وبموجب هذا التحليل يتم تقسيم البيانات إلى مجاميع (عناقيد) لها معنى ودلالات علمية. ويكون التحليل العنقودي في أحيان كثيرة خطوة للدخول في تحليل آخر يحمل أهدافاً تعتمد على تصنيف البيانات إلى عناقيد تحمل أهمية بحثية لدى الباحث، مثال ذلك تلخيص البيانات Data Summarization. وبغض النظر عن كونه بداية لتحليل لاحق أو تصنيف للبيانات، فإن التحليل العنقودي يدخل ضمن استخدامات علمية عديدة وفي مختلف الميادين والتخصصات. ويتضمن مفهوم هذا التحليل معاني عديدة منها أن المفردة أو المشاهدة ضمن عنقود معين تحمل خصائص قريبة لخصائص المشاهدات أو البيانات ضمن العنقود المنتمي إليه، في حين تختلف تلك الخصائص للمشاهدات مع خصائص المشاهدات أو البيانات في العناقيد الأخرى. [يونس والدباغ، 2011]

والتحليل العنقودي هو عبارة عن تجمع من وحدات أو أفراد (Individuals) المجتمع من أجل اكتشاف هيكلية البيانات. إذ تكون الأفراد داخل المجموعة (group) الواحدة قريبة أو متشابهة من بعضها البعض، ولكنها

تختلف عن الأفراد في المجموعات الأخرى. إذ إن نتائج التحليل العنقودي تقدم هيكلية معروفة والتي يمكن أن تستخدم لوضع فرضيات لتفسير البيانات المشاهدة. [Webb, 2002]

ويعتمد التحليل العنقودي على أساليب رياضية وخوارزميات في عملية التصنيف لها القدرة في تحديد مدى التجانس بين مفردات التصنيف أو مدى التناظر بين خصائصها وعندئذ تعطي فرصة للباحث في تفسيرها ضمن قوانين التحليل العنقودي وأساليبه. [يونس والدباغ، 2011]

### *K-Means Algorithm*

### **1-4 خوارزمية K-Means:**

فُدم التعميم الأول لخوارزمية K-Means من قبل العالم Dunn في عام 1974، وهي واحدة من أبسط خوارزميات التعلم بدون إشراف التي تساهم في حل مسائل التحليل العنقودي. إن الهدف الأساسي من هذه الخوارزمية هو تجزئة البيانات إلى K من العناقيد بحيث يتم تصغير مجموع المربعات داخل المجموعة الواحدة Sum of Squares Within-Group.

إن أبسط صيغة لخوارزمية K-Means تعتمد بشكل أساسي على إجراءين متعاقبين، الأول هو تخصيص أو تحديد الكائنات (objects) في مجاميع، إذ يخصص الكائن (object) في المجموعة التي وسطها الحسابي قريب من المسافة الاقليدية. أما الإجراء الثاني فهو حساب متوسطات المجموعات الجديدة بالاعتماد على التخصيصات. وتنتهي العملية عند عدم وجود أي انتقال للكائن إلى مجموعة أخرى وبذلك سوف يقل مجموع المربعات داخل المجموعة الواحدة.

تسمى هذه الخوارزمية أيضاً بخوارزمية C-Means أو خوارزمية الانتقال التكراري Iterative Relocation أو خوارزمية Basic ISODATA. [Webb,2002]. [Abbas and Sadeq, 2014]

### *Fuzzy Clustering*

### **2-4 العنقدة المضيبة:**

تعد خوارزمية العنقدة المضيبة إحدى الخوارزميات المستخدمة بشكل واسع في التحليل العنقودي. وقد اقترحت نظرية التضييب أول مرة من قبل العالم Zadeh في عام 1965 وأعطى فكرة عن اللاتأكدية uncertainty بالانتماء التي وصفها من خلال دالة العضوية Membership Function.

اقترحت تطبيقات نظرية المجموعة المضيبة في التحليل العنقودي مبكراً من خلال عمل كل من الباحثين Belman و Zadeh و Ruspini، إذ فتحت أعمالهم خطوة لباب العنقدة المضيبة. لقد أصبح دمج المنطق المضيبي مع تقنيات تنقيب البيانات واحدة من المكونات الرئيسية للحوسبة في معالجة التحديات التي تفرضها مجموعات كبيرة من البيانات الطبيعية. إن الفكرة الأساسية في العنقدة المضيبة هو تقسيم غير وحيد non-unique partitioning للبيانات إلى مجموعته من العناقيد، ومن نقاط البيانات يتم تعيين قيم العضوية لكل من العناقيد إذ تسمح خوارزمية العنقدة المضيبة للعناقيد بالنمو لتأخذ شكلها الطبيعي.

[Suganya and Shanthi,2012]

وغالبا ما تسمى الأجزاء المضيبة بـ Fuzzy K-Partitions حيث أن K تشير إلى عدد الأصناف المضيبة في الجزء. هناك طريقتين أساسيتين للعقدة المضيبة الأولى تستند على تقنية Fuzzy K-Partitions وتسمى طريقة Fuzzy K-Means Clustering. والطريقة الأخرى تقوم على أساس علاقات التكافؤ المضيبة وتسمى طريقة التعنقد الهرمية القائمة على علاقة التكافؤ المضيبة Fuzzy Equivalence Relation-Based Hierarchical Clustering Method [Klir and Yuan,1995].

### 3-4 خوارزمية K-Means المضيبة: Fuzzy K-Means Algorithm

تسمى هذه الخوارزمية أيضاً بـ (FCM) Fuzzy C-Means والتي قدمت من قبل العالم Bezdek في عام 1981. وهي خوارزمية عنقدة بدون إشراف يتم تطبيقها على مجموعة واسعة من المسائل المرتبطة بتحليل السمات المتمثلة بالعقدة والتصنيف. يتم تطبيق خوارزمية FCM على نطاق واسع في كل من الهندسة الزراعية، علم الفلك، الكيمياء، الجيولوجيا، تحليل الصور وفي التشخيص الطبي. ويتم استخدام هذه الخوارزمية في التحليل بالاعتماد على المسافة بين النقاط المختلفة للبيانات المدخلة. إذ يتم تشكيل العناقيد وفقاً للمسافة بين نقاط البيانات ومراكز العنقود.

إن خوارزمية FCM هي أسلوب عنقدة البيانات التي يتم فيها تجميع مجموعة من البيانات في  $n$  من العناقيد إذ أن لكل نقطة من البيانات في مجموعة البيانات الخاصة بكل عنقود سوف يكون لها درجة عالية من الانتماء (أو الارتباط) بذلك العنقود، أما نقاط البيانات الأخرى التي تقع بعيداً عن مركز العنقود سوف يكون لها درجة انتماء منخفضة لذلك العنقود. كما إنها تعد خوارزمية تكرارية تقوم بنقل مراكز العنقود إلى الموقع المناسب ضمن مجموعة البيانات، ولهذا فإن هذه الخوارزمية مشابهة بشكل أساسي لخوارزمية K-Means وتسلك سلوك مماثل لها. [Ghosh and Dubey, 2013]

إن خوارزمية K-Means لديها خاصية أن كل كائن ينتمي إلى مجموعة واحدة فقط. أما في خوارزمية العنقدة المضيبة فإن الفكرة الأساسية هي السماح للأنماط بالانتماء إلى كافة العناقيد بدرجات عضوية مختلفة. وتحاول خوارزمية FCM إيجاد حل للمعلمة  $y_{ji}$ ، إذ أن  $(i = 1, 2, \dots, n; j = 1, 2, \dots, g)$  التي تصغر الدالة الآتية:

$$J_r = \sum_{i=1}^n \sum_{j=1}^g y_{ji} |x_i - m_j|^2 \quad \dots (3)$$

والخاضعة للقيود التالية:

$$\sum y_{ji} = 1, \quad 1 \leq i \leq n, y_{ji} \geq 0, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, g$$

إذ إن المعلمة  $y_{ji}$  تمثل درجة الارتباط أو دالة العضوية للنمط  $i$  في المجموعة  $j$ . وفي المعادلة (3) فإن  $r$  حد ثابت يسمى دليل الأهمية Weighting Exponent التي تسيطر على الضبابية في نتائج العناقيد أما  $m_j$  فتمثل مركز المجموعة  $j$ :

$$m_j = \frac{\sum_{i=1}^n y_{ji}^r x_i}{\sum_{i=1}^n y_{ji}^r} \quad \dots (4)$$

إن تصغير المعادلة (3) سوف يعطينا قيم  $y_{ji}$  التي تكون إما 0 أو 1. أما الخطوات الأساسية لهذه الخوارزمية فتكون على النحو الآتي كما وصفها العالم Bezdek في عام 1981 [Webb,2002]:

1- اختيار قيمة  $r$  بحيث  $1 < r < \infty$  ، وتهيئة قيم دوال العضوية  $y_{ji}$  حيث أن  $(i = 1, 2, \dots, n; j = 1, 2, \dots, g)$ .

2- حساب مراكز العنقود  $m_j$  حيث أن  $j = 1, 2, \dots, g$  من المعادلة 4.

3- حساب المسافة  $d_{ij}$  من المعادلة التالية عندما:  $(i = 1, 2, \dots, n; j = 1, 2, \dots, g)$

$$d_{ij} = |x_i - m_j|$$

4- حساب دالة العضوية من المعادلة التالية:

$$y_{ji} = \frac{1}{\sum_{k=1}^g \left( \frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{r-1}}}$$

5- إذا لم يتحقق التقارب اذهب إلى الخطوة (2) إذ أن معيار التوقف يعتمد على التغير في قيم دالة العضوية  $y_{ji}$  أو على التغير في قيم دالة الهدف  $J_r$ .

## (5) الجانب التطبيقي: Application

لتحقيق هدف البحث ألا وهو اقتراح نموذج ماركوفي مصاغ على أساس مبادئ العنقدة المضببة ومقارنته مع نموذج ماركوفي مصاغ على أساس العنقدة غير المضببة، وذلك بالاعتماد على خوارزمية K-Means للعنقدة، تم تناول جانباً تطبيقياً متمثلاً ببعض محتويات الفاكهة والخضار التي تمثل نسبة البروتين والدهنيات والسكريات والألياف والماء والسرعات الحرارية، وتم الحصول على هذه البيانات من مجلة العراق الدوائية الصادرة عن اتحاد الصيادلة العرب عام 2011 ، وتضم البيانات 36 نوع من الفواكه والخضراوات. إذا تم إجراء التحليل العنقودي لهذه البيانات باستخدام خوارزميتين: الأولى هي خوارزمية K-means والثانية هي خوارزمية Fuzzy C-means وذلك لتقسيم البيانات إلى مجاميع بالاعتماد على السرعات الحرارية كمداخل لسلسلة ماركوف وكما موضح كالتالي:

### (1-5) النموذج الماركوفي باستخدام خوارزمية K-Means:

تم إجراء التحليل العنقودي للبيانات باستخدام خوارزمية K-Means وكانت نتائج تنفيذ هذه الخوارزمية على البيانات هو تصنيفها إلى مجموعتين أو عنقودين وكما موضحه في الجدول الآتي:

جدول (1): نتائج تصنيف البيانات باستخدام خوارزمية K-Means.

التصنيف بخوارزمية K-Means	السرعات الحرارية	نسبة الماء	نسبة الألياف	نسبة السكريات	نسبة الدهون	نسبة البروتين	اسم الفاكهة أو الخضار
1	30	92	2.6	7	0.4	0.6	الفراولة
2	71	81	1.2	18	0.6	0.7	العنب
2	55	84	4.4	14	0.2	1.4	الكشمش
2	51	86	4.6	13	0.4	0.7	التوت البري
2	56	85	2.3	14	0.4	0.7	الأس
2	49	87	4.7	12	0.6	0.9	توت العليق
2	50	84	2.2	15	0.4	0.2	التفاح
2	55	85	1.6	13	0.6	0.8	الخوخ
1	43	88	1.6	11	0.1	0.7	الدراق
2	59	84	2.6	15	0.4	0.4	العجاص
2	58	84	1.7	15	0.1	0.4	السفرجل
1	48	86	0.6	11	0.4	1.4	المشمش
1	0	81	1.6	17	1	1.2	الكرز
1	47	87	0.4	12	0.2	0.9	البرتقال
1	37	89	2.1	9.3	0.3	1	الليمون الحامض
1	37	89	0.4	9.6	0	0.7	الكريب فروت
1	35	89	0	8	0	0.5	البطيخ
1	0	90	0	7	0.4	0.6	الرقى
1	24	95	1.8	5	0.2	2	القرنبيط
2	79	92	1.5	18	0.1	2.1	البطاطا
2	105	28.3	2.5	24.3	0.3	1.6	البطاطا الحلوة
1	24	73	1.8	5.4	0.2	1.2	الملفوف
1	21	93	2.6	3.5	0.3	2.9	السبانغ
1	13	91.6	1.2	4.6	0.3	0.8	البندورة
1	27	93.8	0.7	2.9	0.1	0.5	الخيار
1	20	96	2	6.4	0.2	0.9	الفليفلة
1	27	93.7	1.6	4.3	0.3	0.9	القرع
1	43	92	1.5	0	6.3	1.2	الباذنجان
1	27	87.8	3.2	6.6	0.1	0.9	الجزر
1	27	89	2.9	5.5	0.2	2.6	الثوندر
1	17	95	2.2	3.6	0.5	0.6	الفجل
1	15	95	1.5	3.8	0	0.8	الكرفس
1	19	92.7	1.6	3.7	0.2	1.8	السلق
2	60	93	1.8	14	0.3	1.5	الكراث

2	144	59	0.14	3	0.1	0.6	الثوم
1	38	89.7	1.6	8.6	0.2	1.2	البصل

إذ أن العمود الأخير من الجدول يمثل نوع التصنيف الذي وقعت فيه المشاهدة، وهما الصنفين: { 2، 1 }. والجدول التالي يمثل ملخص لكل صنف أو عنقود تم الحصول عليه باستخدام خوارزمية K-Means.

جدول (2): المعلومات الخاصة بالعناقيد التي تم الحصول عليها باستخدام خوارزمية K-Means.

Clusters	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster 1	23	90.215	1.672	5.820
Cluster 2	13	72.885	2.034	4.897

من خلال النتائج السابقة تم ملاحظة انه تم تقسيم البيانات إلى عنقودين، ضم العنقود الأول (23) مشاهدة أما العنقود الثاني فقد ضم (13) مشاهدة. ومصفوفة المسافة بين مراكز العنقودين فتكون على وفق الشكل الآتي:

$$D = \begin{matrix} & \text{cluster 1} & \text{cluster 2} \\ \text{cluster 1} & [0.0000 & 2.3763] \\ \text{cluster 2} & [2.3763 & 0.0000] \end{matrix}$$

من خلال تنفيذ خوارزمية K-Means على البيانات قيد الدراسة تم الحصول على مجموعتين من الفواكه والخضراوات، المجموعة الأولى هي الفواكه والخضراوات منخفضة السرعات الحرارية والمتمثلة بالعنقود الأول، إذ يكون المدى لهذه المجموعة هو (0-48) سعره حرارية. أما المجموعة الثانية فهي مرتفعة السرعات الحرارية والمتمثلة بالعنقود الثاني إذ يكون المدى لهذه المجموعة هو (49-144) سعره حرارية.

وكمدخلات للنموذج الماركوفي فقد تم استخدام العناقيد التي تم الحصول عليها كفضاء حالة للنموذج الماركوفي والمتمثل بالفواكه والخضراوات منخفضة السرعات الحرارية، مرتفعة السرعات الحرارية، إذ تم إيجاد التكرارات من خلال الانتقالات التي تحدث بين الفواكه والخضراوات من الحالة منخفضة السرعات الحرارية إلى الحالة منخفضة السرعات الحرارية، ومن الحالة منخفضة السرعات الحرارية إلى مرتفعة السرعات الحرارية، وكذلك من مرتفعة السرعات الحرارية إلى منخفضة السرعات الحرارية، ومن حالة مرتفعة السرعات الحرارية إلى مرتفعة السرعات الحرارية، وقد تم وضع التكرارات في جدول مماثل للمصفوفة الانتقالية وكما مبين أدناه:

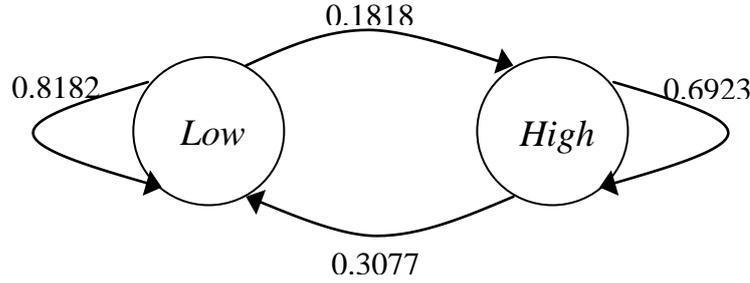
## جدول (3): مصفوفة التكرارات للبيانات الناتجة من تصنيف خوارزمية K-Means.

الحالة الحالية الحالة السابقة	منخفضة السرعات الحرارية (Low)	مرتفعة السرعات الحرارية (High)	المجموع
منخفضة السرعات الحرارية (Low)	18	4	22
مرتفعة السرعات الحرارية (High)	4	9	13

وحسب تعريف التكرار النسبي للاحتتمالية تم الحصول على مقدرات للاحتتمالات الانتقالية من خلال تقسيم كل تكرار في الخلايا الأربعة على المجموع الهامشي لكل صف، وكذلك تم إيجاد المصفوفة الانتقالية باستخدام اللغة البرمجية *Matlab V.7.12* كالآتي:

$$P = \begin{matrix} & \begin{matrix} Low & High \end{matrix} \\ \begin{matrix} Low \\ High \end{matrix} & \begin{bmatrix} 0.8182 & 0.1818 \\ 0.3077 & 0.6923 \end{bmatrix} \end{matrix}$$

أما المخطط الانتقالي لهذه الحالات فقد تم توضيحه في الشكل الآتي:



## الشكل (1): المخطط الانتقالي للمصفوفة الانتقالية للتصنيف باستخدام خوارزمية K-Means

من المخطط الانتقالي في الشكل رقم (1) تم توضيح انتقال صفة السرعات الحرارية ( المنخفضة والمرتفعة) بين الفواكه والخضراوات، كما تم ملاحظة أن السلسلة غير قابلة للتجزئة Irreducible، وأن جميع الحالات تكون لها نفس الخصائص، إذ تم ملاحظة أنها ذات حالة عوددة موجبة Positive Recurrent وغير دورية Aperiodic، وعليه وبما أنها غير قابلة للتجزئة وذات عودة موجبة وغير دورية فإنها تكون ثبوتية Ergodic، وبذلك يمكن إيجاد التوزيع المراوح Stationary Distribution للحالتين وذلك من خلال استخدام اللغة البرمجية *Matlab V.7.12* وكما مبين أدناه:

$$\pi_{Low} = 0.6286 \quad , \quad \pi_{High} = 0.3714$$

إذ أن مصفوفة التوزيع المراوح استقرت عند التكرار  $n = 16$ . ومن نسب التوزيع المراوح أعلاه تم ملاحظة أن نسبة استقرار الفواكه والخضراوات على صفة انخفاض السرعات الحرارية هو 62.82%، في حين أن نسبة استقرار الفواكه والخضراوات على صفة ارتفاع السرعات الحرارية هو 37.14%.

## 5-2) النموذج الماركوفي باستخدام خوارزمية Fuzzy C-Means:

في هذه الفقرة تم إجراء التحليل العنقودي المضرب للبيانات باستخدام خوارزمية Fuzzy C-Means وكانت نتائج تنفيذ هذه الخوارزمية على البيانات هو تصنيفها إلى مجموعتين أو عنقودين، وتم الحصول على قيم دوال العضوية لكل مشاهدة بالاستعانة باللغة البرمجية V.7.12 *Matlab* ومن خلال الدالة *fcm*، إذ توقفت عملية العنقدة بعد 19 تكرار وكانت عندها قيمة دالة الهدف *obj. fcn* مساوية لـ (12789.497123) وكما مبين في النتائج الآتية:

Iteration count = 1، *obj. fcn* = 17451.013189  
 Iteration count = 2، *obj. fcn* = 15346.428979  
 Iteration count = 3، *obj. fcn* = 13471.352894  
 Iteration count = 4، *obj. fcn* = 12876.073387  
 Iteration count = 5، *obj. fcn* = 12801.196729  
 Iteration count = 6، *obj. fcn* = 12792.054738  
 Iteration count = 7، *obj. fcn* = 12790.259697  
 Iteration count = 8، *obj. fcn* = 12789.747574  
 Iteration count = 9، *obj. fcn* = 12789.581489  
 Iteration count = 10، *obj. fcn* = 12789.525766  
 Iteration count = 11، *obj. fcn* = 12789.506879  
 Iteration count = 12، *obj. fcn* = 12789.500450  
 Iteration count = 13، *obj. fcn* = 12789.498258  
 Iteration count = 14، *obj. fcn* = 12789.497509  
 Iteration count = 15، *obj. fcn* = 12789.497254  
 Iteration count = 16، *obj. fcn* = 12789.497166  
 Iteration count = 17، *obj. fcn* = 12789.497136  
 Iteration count = 18، *obj. fcn* = 12789.497126  
 Iteration count = 19، *obj. fcn* = 12789.497123

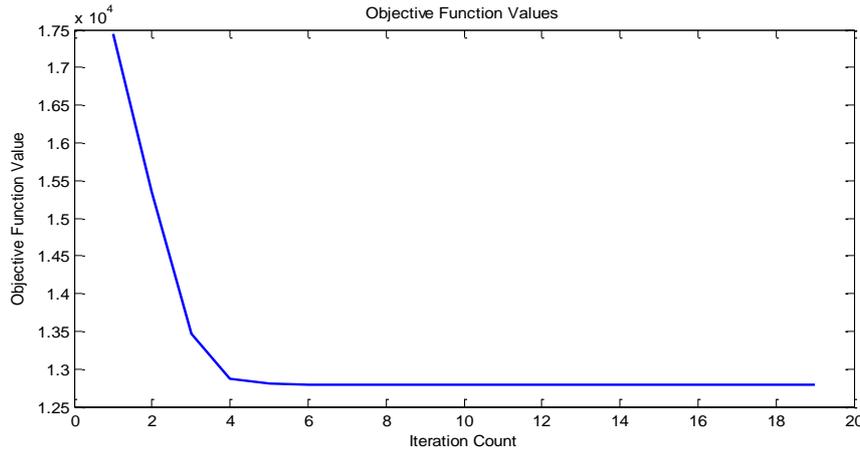
ونائج التصنيف لخوارزمية FCM ودوال العضوية لكل مشاهدة موضحة في الجدول أدناه:

### جدول (4): نتائج تصنيف البيانات باستخدام خوارزمية Fuzzy C-Means .

التصنيف بخوارزمية FCM	السيطرة الضبابية	اسم الفاكهة أو الخضار
1	0.9793	الفاولة
2	0.9592	العنب
2	0.9430	الكشمش
2	0.8330	التوت البري
2	0.9582	الأس
2	0.7463	توت العليق
2	0.8242	التفاح
2	0.9408	الخوخ
1	0.5523	الدراق
2	0.9892	العنجااص
2	0.9824	السفرجل

2	0.7069	المشمش
1	0.8195	الكرز
2	0.6612	البرتقال
1	0.8323	الليمون الحامض
1	0.8287	الكريب فروت
1	0.8954	البطيخ
1	0.8577	الرقبي
1	0.9830	القرنبيط
2	0.8847	البطاطا
2	0.6814	البطاطا الحلوة
1	0.8464	الملفوف
1	0.9790	السبانغ
1	0.9403	البندورة
1	0.9811	الخيار
1	0.9679	الفليفلة
1	0.9871	القرع
1	0.6446	الباذنجان
1	0.9928	الجزر
1	0.9947	الشوندر
1	0.9560	الفجل
1	0.9456	الكرفس
1	0.9730	السلق
2	0.9163	الكرات
2	0.6702	الثوم
1	0.8067	البصل

من الجدول (4) تم ملاحظة انه تم تقسيم البيانات إلى مجموعتين، ضمت المجموعة الأولى (21) مشاهدة أما المجموعة الثانية فقد ضمت (15) مشاهدة. والشكل الآتي يوضح العلاقة بين قيم دالة الهدف وعدد التكرارات.



### الشكل (2): العلاقة بين دالة الهدف وعدد التكرارات الناتجة عن تنفيذ خوارزمية Fuzzy C-Means

من خلال تنفيذ خوارزمية Fuzzy C-Means على البيانات قيد الدراسة تم الحصول على مجموعتين من الفواكه والخضراوات، المجموعة الأولى هي الفواكه والخضراوات منخفضة السرعات الحرارية والمتمثلة بالعنقود الأول، إذ يكون المدى لهذه المجموعة هو (0-45) سعره حرارية. أما المجموعة الثانية فهي مرتفعة السرعات الحرارية والمتمثلة بالعنقود الثاني إذ يكون المدى لهذه المجموعة هو (46-144) سعره حرارية.

وكمدخلات للنموذج الماركوفي فقد تم استخدام العناقيد التي تم الحصول عليها كفضاء حالة للنموذج الماركوفي والمتمثل بالفواكه والخضراوات {منخفضة السرعات الحرارية، مرتفعة السرعات الحرارية}، وقد تم إيجاد التكرارات للحالتين أي إيجاد عدد الانتقالات بين الحالات وتم وضعها في جدول مماثل للمصفوفة الانتقالية وكما مبين أدناه:

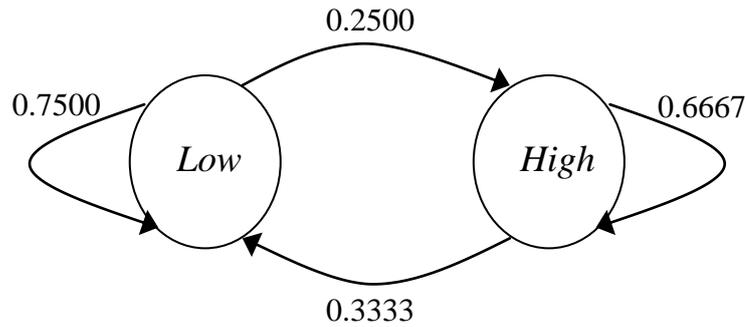
### جدول (5): مصفوفة التكرارات للبيانات الناتجة من تصنيف خوارزمية Fuzzy C-Means

الحالة الحالية \ الحالة السابقة	منخفضة السرعات الحرارية (Low)	مرتفعة السرعات الحرارية (High)	المجموع
منخفضة السرعات الحرارية (Low)	15	5	20
مرتفعة السرعات الحرارية (High)	5	10	15

وحسب تعريف التكرار النسبي للاحتتمالية تم الحصول على مقدرات للاحتتمالات الانتقالية، وكذلك تم إيجاد المصفوفة الانتقالية باستخدام اللغة البرمجية Matlab V.7.12 للنموذج الماركوفي المضرب كآتي:

$$P = \begin{matrix} & \text{High} & \text{Low} \\ \text{Low} & 0.7500 & 0.2500 \\ \text{High} & 0.3333 & 0.6667 \end{matrix}$$

أما المخطط الانتقالي لهذه المصفوفة فيكون بالشكل الآتي:



### الشكل (3): المخطط الانتقالي للمصفوفة الانتقالية للتصنيف باستخدام خوارزمية Fuzzy C-Means

من الشكل (3) تم ملاحظة أن السلسلة غير قابلة للتجزئة، وعليه فإن جميع الحالات تكون لها نفس الخصائص إذ تبين أنها حالة عودة موجبة وغير دورية، وبذلك فإن السلسلة تكون ثبوتية، وبذلك يمكن إيجاد التوزيع المراوح للحالتين وذلك من خلال استخدام اللغة البرمجية *Matlab V.7.12* وكما مبين أدناه:

$$\pi_{F_{High}} = 0.4286 \pi_{F_{Low}} = 0.5714$$

وقد استقرت مصفوفة التوزيع المراوح عند التكرار  $n = 12$ . ومن نسب التوزيع المراوح أعلاه تم ملاحظة أن نسبة استقرار الفواكه والخضراوات على صفة انخفاض السرعات الحرارية بعد استخدام العقدة المضببة للنموذج الماركوفي هو 57.14%، في حين أن نسبة استقرار الفواكه والخضراوات على صفة ارتفاع السرعات الحرارية هو 42.86%.

## Conclusions

## (6) الاستنتاجات:

- أ- مما تقدم تم التوصل إلى انه يمكن الدمج بين النموذج الماركوفي والتحليل العنقودي في تصنيف الحالات، وذلك من خلال اقتراح نماذج ماركوفيه مصاغة على أساس مبادئ العنقدة المضبية، أو العنقدة غير المضبية.
- ب- تم ملاحظة أن مصفوفة التوزيع الماروح للحالات للنموذج الماركوفي المصاغ على أساس العنقدة المضبية استقرت بشكل أسرع من النموذج الماركوفي المصاغ على أساس العنقدة غير المضبية.
- ج- من المخطط الانتقالي لانتقال صفة السرعات الحرارية (المنخفضة والمرتفعة) بين الفواكه والخضراوات قيد الدراسة، تم التوصل إلى أن سلسلة الانتقالات غير قابلة للتجزئة Irreducible، وان جميع الحالات لها نفس الخصائص، وذات حالة عودة موجبة Positive Recurrent وغير دورية Aperiodic، وعليه فأنها تكون ثبوتية Ergodic.
- د- من نسب التوزيع الماروح للنموذج الماركوفي المصاغ على أساس العنقدة غير المضبية تم ملاحظة أن نسبة استقرار الفواكه والخضراوات قيد الدراسة على صفة انخفاض السرعات الحرارية أكبر من استقرارها على صفة ارتفاع السرعات الحرارية، مما يدل على إن الفواكه والخضراوات قيد الدراسة ليس فيها سرعات حرارية عالية. في حين أن نسب التوزيع الماروح للنموذج الماركوفي المصاغ على أساس العنقدة المضبية للحالتين كانت متقاربة تقريبا.
- هـ- من خلال ملاحظة نتائج تصنيف البيانات باستخدام خوارزمية Fuzzy C-Means جدول (4) نلاحظ أن جميع قيم دوال العضوية تقترب من الواحد مما يدل على إن درجة ارتباط أو انتماء العنصر بالعنقود أو المجموعة التي ينتمي إليها كبيرة.

## References

## المصادر:

1. احمد، ريكان عبد العزيز، (2008). " سلاسل ماركوف بين النظرية والتطبيق في المجال الاقتصادي او المالي او الاداري " ، مجلة تنمية الراءدين، المجلد30، العدد 92.
2. الخياط، باسل يونس ذنون ، (2011). " النمذجة الماركوفية مع تطبيقات عملية "، دار الكتب والنشر- الموصل، مطبعة جامعة الموصل.
3. مجلة العراق الدوائية الصادرة عن اتحاد الصيادلة العرب، العدد 10، تشرين الثاني 2011.
4. يوسف، مناهل عبد الكريم، (2011). " استخدام أحدى طرق العقدة والمنطق المضرب في تصنيف صور الأنسجة" ، مجلة تكريت للعلوم الصرفة، المجلد 16، العدد 3.
5. يونس، طارق شريف و رائد عبد القادر الدباغ، (2011). " التحليل الإحصائي مفاهيم- منهجية تطبيقات استخدام البرمجية *Minitab*"، جامعة العلوم التطبيقية، مملكة البحرين.
6. Abbas A. Khudhair and Sadeq A. Tariq, (2014). "*Database Clustering using Intelligent Techniques*", Journal of Al-Nahrain University, Vol. 17, No.3, 195-203.
7. Fathi Vajargah, B., Gharehdaghi, M., (2012). "*Ergodicity of Fuzzy Markov Chains Based on Simulation Using Halton Sequences*". The journal of Mathematics and Computer Science, 380-385.
8. Fathi Vajargah, B., Gharehdaghi, M., (2014). "*Ergodicity of Fuzzy Markov Chains Based on Simulation Using Sequences*". The journal of Mathematics and Computer Science, Vol.11, 159-165.
9. Fathi Vajargah, B., Masoumi, H., (2012)."*Study on Behavior of Fuzzy Markov Chains*". Advances in Computer Science and its Application (ACSA), Vol 2, No.2, 2166-2924.
- 10.Ghosh Soumi and Dubey Sanjay Kumar, (2013). "*Comparative Analysis of K-Means and Fuzzy C-Means Algorithms*", IJACSA, Vol.4, No.4.
- 11.Klir George and Yuan Bo, (1995). "*Fuzzy Sets and Fuzzy Logic Theory and Applications*", Published by Prentice Hall PTR.
- 12.Marius Iosfescu, (1980). "*Finite Markov Process and Their Applications*", John Wiley and Sons .
- 13.Naes Tormod and Mevik Bjorn-Helge, (1999). "*The flexibility of Fuzzy Clustering illustrated by Examples*", Journal of Chemometrics, 13, 435-444.
- 14.Suganya R. and Shanthi R., (2012). "*Fuzzy C-Means Algorithm – A Review*", International Journal of Scientific and Research Publications, Vol 2, 11.
- 15.Webb Andrew R., (2002). "*Statistical Pattern Recognition*", Second Edition, John Wiley and Sons, Ltd.