نموذج ضبابي لتحليل بيانات سلوك العملاء وتصنيفهم في بيئة التجارة الإلكترونية

A fuzzy model for analyzing and classifying customer behavior data in an e-commerce environment.

د. محمّد حجّوز

الملخص

نظراً لكثرة مواقع التواصل الاجتماعي ومواقع الشراء الإلكتروني وما ينتج عنه من بيانات ضخمة، كانت الحاجة ماسة لاعتماد مجال تعلم الآلي والمعالجة الألية للغات الطبيعية، وذلك من أجل تنظيم ما يتم تداوله على هذه المواقع ومعالجتها بأقل وقت وجهد ممكنين.

وللحاجة لدراسة تحليل المشاعر والعواطف والأراء بهدف بناء نظام يقوم على تحليل ما يكتبه الناس للتعبير عن أرائهم بصور نصية أو من خلال التغريدات أو من خلال الرسوم التعبيرية.

ومن خلال دراسات سابقة تقوم على تقييم العملاء كما في الأنظمة الناصحة بالتجارة الالكترونية. وتحليل سلة المشتريات واختيار العميل لأصناف المنتجات ومسار التنقل بينها، والأبحاث التي تركز على التقنيات والخوارزميات المستخدمة في عملية التنبؤ بشكل أكبر وذلك في محاولة لتحسين دقة التنبؤ.

قمنا بهذا البحث بتحديد السلوكيات المعقدة للعميل من خلال السماح بالانتماء الجزئي لمجموعات متعددة، مما يعكس الواقع بشكل أفضل (حيث يمكن أن يكون للعميل أكثر من فئة سلوكية).

وتم استخدام التجميع الضبابي (خوارزمية FCM) مع عامل زمن التصفح والحسميات وحساب المعامل Feature الجديد الذي يعتمد على عدة عوامل، مع الأخذ بعين الاعتبار ولاء العميل وفهم سلوكه والتنبؤ به لمعرفة العملاء الذين يقومون بعملية التصفح فقط دون الشراء أو الذين يقومون بعملية الشراء فورا.

الكلمات المفتاحية:

Fuzzy-C-Means التجميع الضبابي، Silhouette Score مقياس سوليت، Partition Coefficient معامل التقسيم.

Abstract

Given the proliferation of social media and e-commerce platforms, and the resulting massive amounts of data, there is a pressing need to adopt machine learning and natural language processing (NLP) to organize and process the content shared on these platforms with minimal time and effort.

There is also a need to study sentiment analysis and opinion analysis to build a system that analyzes what people write to express their opinions, whether through text, tweets, or emojis.

Previous studies have focused on customer evaluation, as seen in e-commerce advisory systems. These studies analyze shopping carts, customer product selection, and navigation patterns, and research has further emphasized the technologies and algorithms used in prediction processes to improve prediction accuracy. In this research, we identified the complex behaviors of customers by allowing partial affiliation to multiple groups, which better reflects reality (where the customer is in more than one behavioral category). We used fuzzy clustering (FCM algorithm) and took into account the browsing time factor, discounts, and the calculation of the new Feature parameter, which depends on several factors, taking into account customer loyalty, understanding their behavior, and predicting it to know which customers only browse without buying or who make a purchase immediately.

Key Words:

Fuzzy-C-Means (FCM), Silhouette Score, Partition Coefficient

1 مقدمة:

مع ازدياد حجم البيانات على كافة الأصعدة والوصول بها إلى حجم البيانات الضخمة، ومع تطور أدوات الشبكة وظهور مواقع التواصل الاجتماعي والمنتديات ومواقع البيع والشراء الإلكتروني وغيرها، ظهرت الحاجة ماسة لاعتماد مجال تعلم الآلي والمعالجة الألية للغات الطبيعية، من أجل تنظيم ما يتم تداوله على هذه المواقع ومعالجتها بأقل وقت وجهد. وظهرت الحاجة أيضاً إلى دراسة تحليل المشاعر والعواطف والأراء بهدف بناء نظام يقوم على تحليل ما يكتبه الناس للتعبير عن أرائهم بصور نصية أو من خلال التغريدات أو من خلال الرسوم التعبيرية.

ساهمت علوم الذكاء الإصطناعي والتعلم الآلي والتعلم العميق في حل تطبيقات عديدة في مجال معالجة اللغات الطبيعية على مستوى كبير من الاهتمام للمساهمة في خدمة البشرية وفي التعامل مع وسائل التكنولوجيا ومعرفة التوجهات وسلوك الأفراد.

2- مشكلة البحث

تتمثل المشكلة العلمية في أن بيانات سلوك العملاء في بيئة التجارة الإلكترونية تتسم بالتعقيد والتنوع وعدم اليقين، نتيجة اختلاف أنماط التصفح والشراء والتفاعل مع المنتجات، إضافة إلى وجود تداخل بين خصائص العملاء وعدم وجود حدود واضحة تفصل بين الفئات السلوكية المختلفة. وهذا يجعل تصنيف العملاء باستخدام الأساليب التقليدية القائمة على الحدود الحادة أقل قدرة على تمثيل السلوك الحقيقي للعميل. ومن هنا تظهر الحاجة إلى تطوير نموذج يعتمد على المنطق الضبابي، بما يسمح بتمثيل درجة انتماء العميل إلى أكثر من نمط أو فئة سلوكية، وتحليل سلوكه بصورة أكثر مرونة ودقة. تواجه الشركات التجارية التحديات الجمّة في استهداف العملاء وجذبهم ليكونوا زبائن دائمين ومخلصين، وهذا يتطلب تحليل سلوك العميل المتبع قبل الوصول إلى عملية الشراء لكسب رضا العملاء وفهم احتياجاته وتوفير خدمات مخصصة لهم.

لهذا لا بد من إجراء تحليل فترات الدخول ومعدل الارتداد ومعرفة الصفحات المفضلة للزوار وفهم سلوك الزائر وتجربته على الموقع وهذا ما يسمى (تحليلات الويب) أو تحليل السلوك على الانترنت من خلال تحليل التفاعلات والاهتمامات والتعليقات والتغريدات على توتير (**X**).

3- أهمية البحث:

إن عملية تحليل بيانات سلوك العميل تُحسّن من مستوى الأعمال، ويمكن أن تكشف عن معلومات مخفية مثل تفضيلات العملاء والصفحات الشائعة على موقع الويب والوقت الذي يقضيه العملاء في التصفح وتعليقات العملاء والتفاعل مع نماذج مواقع الويب وهذا يُمكّن الشركات من الاستجابة بكفاءة لاحتياجات العملاء وزيادة مستوى الرضا لديهم.

4- هدف البحث:

تسعى الشركات إلى جذب العملاء لشراء منتجاتها من خلال مواقع التجارة الالكترونية من خلال تحديد (معرفة) السلوك المتبع للعميل قبل عملية الشراء باستخدام تقنيات تحليل البيانات والتعلم الالي. ويهدف هذا البحث إلى معرفة المجموعات السلوكية الرئيسية للعملاء التي تكشف عنها FCM، ومعالجة المجموعات المتداخلة، والمعالجة العلمية للنتائج واتخاذ القرارات التسويقية.

5- الدراسات السابقة:

العديد من الأبحاث والدراسات ساهمت بتقديم كم هائل من المعلومات والخبرات التي تُمكّن الباحثين من الحصول على معرفة شاملة حول مواضيع مُحدَّدة، ومن هذه الدراسات نذكر بعض الدراسات على النحو الأتي:

5-1 استخدام الذكاء الاصطناعي في تحسين إدارة علاقات العملاء:[1]

Customer relationship management through artificial intelligence technologies

نشر الباحثون Zahra Sabila Nugraha و Maulana Asykari Muhammad و Bayu Waspodo هذا البحث قي مجلة جامعة Syarif Hidayatullah في اندوسيا بالعام 2024، وركّزت الدراسة على إدارة علاقات العملاء الاستراتيجية للحفاظ عليهم. واعتمدت على دمج الذكاء الاصطناعي وإدارة علاقات العملاء والبيانات الضخمة لمواجهة تحديات الأعمال الحديثة في عصر التحول الرقمي. تُظهر نتائج هذه الدراسة أن التعلم الخاضع للإشراف لا يزال أكثر تقنيات الذكاء الاصطناعي استخدامًا، بينما شهد التعلم العميق نموًا ملحوظًا في السنوات الأخيرة، مما يشير إلى التحول نحو حلول أكثر تطورًا لإدارة علاقات العملاء، لاسيما في التنبؤ بعزوف العملاء، وتقسيمهم، وتخصيص خدماتهم. لكن لم توضح هذه الدراسة السلوكيات المعقدة للعملاء عندما ينتمون وبشكل جزئي لعدة مجموعات.

2-5 دراسة تحليلية للعوامل المؤثرة على القيمة الدائمة للعميل وفق نموذج RFM (Recency, Frequency, and Monetary):[2]

An analytical study of the factors influencing the customer's future vision according to the RFM model.

قام بهذه الدراسة كل من T Balhaj و S Kwat في معهد Institute of Economic, Commercial and Management Sciences في الجزائر بعام 2023، وتعتمد على إدارة علاقات العملاء بشكل كبير على مقياس القيمة الدائمة للعميل في تقسيم العملاء على أساس ربحهم وتهدف هذه الدراسة إلى اختبار مدى كفاءة نموذج RFM المدمج مع تقنيات التنقيب في البيانات وهذه الدراسة تعطي معلومات حول العوامل التي تتحكم في القيمة الدائمة للعميل. لكن لم توضح هذه الدراسة السلوكيات المعقدة للعملاء عندما ينتمون وبشكل جزئي لعدة مجموعات.

3-5 الذكاء الاصطناعي في إدارة علاقات العملاء: إطار عمل منهجي لتحقيق التكامل الناجح:[3]

Artificial intelligence in customer relationship management: A systematic framework for a successful integration

قام بهذه الدراسة كل من Cristina Ledro و Anna Nosella و Ilaria Dalla Pozza و Andrea Vinelli و Thomas Souverain وتم نشرها في مجلة Journal of Business Research بالعدد 199 بشهر تشرين الأول من العام 2025، وتقدم هذه الدراسة إطارًا منهجيًا لدمج تقنيات الذكاء الاصطناعي في إدارة علاقات العملاء، مع التركيز على الجوانب التنظيمية والتقنية والأخلاقية اللازمة لنجاح هذا التكامل. إلا أن تركيزها ينصب بصورة أساسية على **إطار دمج الذكاء الاصطناعي في إدارة علاقات العملاء**، ولم تتناول بصورة مباشرة مشكلة **التداخل بين الأنماط السلوكية للعملاء عند تحليل بياناتهم وتصنيفهم**، كما لم تبحث في إمكانية تمثيل العميل بدرجات انتماء متفاوتة إلى أكثر من مجموعة سلوكية في الوقت نفسه. وتمثل هذه النقطة فجوة يمكن معالجتها من خلال بحثنا، الذي يقترح استخدام **نموذج ضبابي لتحليل بيانات سلوك العملاء وتصنيفهم**، بما يسمح بالتعامل مع الحدود غير الواضحة بين الفئات السلوكية وتمثيل درجة انتماء العميل إلى كل فئة بصورة أكثر مرونة.

6- التنقيب في البيانات والتعلم الآلي:

يُعد اكتشاف الأنماط والمعرفة الكامنة في البيانات من الممارسات التي رافقت تطور عملية تحليل المعلومات منذ وقت مبكر، إلا أن ظهور الحواسيب وتطور قدراتها الحاسوبية وزيادة إمكانات التخزين أسهما في الانتقال من الأساليب اليدوية إلى أساليب آلية أكثر كفاءة في معالجة كميات كبيرة من البيانات وتحليلها. وقد أدى ذلك إلى تطور مجموعة واسعة من الأساليب والتقنيات التي تهدف إلى استخراج المعرفة ذات القيمة من البيانات واكتشاف العلاقات والأنماط غير الظاهرة فيها. ويُعرف التنقيب في البيانات بأنه مجموعة من الأساليب التي تُستخدم لاكتشاف الأنماط والمعارف المخفية في مجموعات البيانات الكبيرة، وتحويل البيانات الخام إلى معلومات يمكن الاستفادة منها في التحليل واتخاذ القرار.[4]

وتتعدد تطبيقات التنقيب في البيانات تبعًا لطبيعة البيانات والمشكلة المراد تحليلها؛ فمنها التنقيب في النصوص والمشاعر الذي يركز على تحليل الكلمات والعبارات واكتشاف العلاقات الدلالية والأنماط المتشابهة داخل البيانات النصية، ومنها التنقيب في السلاسل الزمنية الذي يهتم بتحليل البيانات المرتبطة بالزمن والاستفادة من الأنماط السابقة في التنبؤ بالقيم أو الأحداث المستقبلية، مثل التنبؤ باتجاهات أسعار الأسهم أو توقعات حالة الطقس. [5][6]

7-النظم الناصحة Recommendation System:

تلعب النظم الناصحة دوراً محورياً في تحسين تجربة المستخدم وزيادة المبيعات في منصات التجارة الالكترونية، ولها دوراً مهماً وهو تقديم النصائح للمساعدة في إيجاد ما نحتاج إليه وتزويد المستخدمين باقتراحات عن العناصر والسلع لكي يتم استخدامها. تعتمد هذه الأنظمة على خوارزميات ذكية لتحليل سلوك المستخدمين وتقديم منتجات أو خدمات تلائم تفضيلاتهم من أشهر هذه الخوارزمية التصفية التعاونية Collaborative Filtering (CF) والتي تعمل على إيجاد تقييمات المستخدمين لمنتجات قد يفضلونها كما أن خوارزميات التعلم الآلي الخاضعة للإشراف التنبُّؤية تشبه النظم الناصحة المعتمدة على التصفية التعاونية CF وذلك عندما تكون طبيعة الهدف عبارة عن قيم عددية مستمرة Continuous، ولتدريبهما نقوم بتقسيم مجموعة معطيات إلى مجموعتين: مجموعة للتدريب والأُخرى لاختبار العيِّنات، ومن ثَمَّ إيجاد البارامترات للنموذج الذي يتم تدريبه بحيث يقدم أفضل النتائج.[7][8]

8-١ Clustering

تعتبر خوارزميات التعلم غير الخاضع للإشراف (Unsupervised Learning) إحدى فروع التعلم الآلي الذي يهدف إلى اكتشاف أنماط أو تراكيب خفية في البيانات دون الاعتماد على تسميات مسبقة. واكتشاف

أنماط في البيانات غير المصنفة. يعتمد الذكاء الاصطناعي بشكل كبير على هذه الخوارزميات لتحويل البيانات الخام إلى رؤى قابلة للتنفيذ سواء في القطاعات الصحية أو التجارية أو التكنولوجية.[8]

9- التجميع الضبابي:

التجميع الضبابي هو إحدى تقنيات التجميع في التعلم الآلي التي تسمح بنقاط البيانات بالانتماء إلى أكثر من مجموعة واحدة بنسب متفاوتة، بدلاً من الانتماء الحصري لمجموعة واحدة كما في K-Means ومن أهم خوارزميات التجميع الضبابي (Fuzzy Cluster –c–Mean).

1- Fuzzy Cluster - c–Mean:

تفيد هذه الخوارزمية بشكل خاص في تحليل سلوك العملاء، حيث يمكن أن يكون للعميل سمات متعددة تجعله ينتمي إلى أكثر من مجموعة سلوكية في نفس الوقت وتعتبر FCM أداة قوية في التعلم غير الخاضع للإشراف، وخاصة في السياقات التي تتطلب مرونة في تحديد انتماء البيانات. والتفاعل مع التطبيقات (مثل الوقت المستغرق، الصفحات التي تمت زيارتها) والتسويق الشخصي مثل إرسال إشعارات مخصصة بناء على درجات العضوية في الكتل.[8]

- **خوارزمية FCM:** تمتلك هذه الخوارزمية المنهجية الآتية:

1-تهيئة أولية: اختيار عشوائي لمراكز الكتل.

2-حساب درجات العضوية: تحدد درجة انتماء كل نقطة بيانات لكل كتلة باستخدام دالة رياضية تعتمد على حساب المسافة بين النقطة ومركز الكتل و دالة الهدف من خلال العلاقة الرياضية:

$$j = \sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}^{m}.\left\|x_i - u_j\right\|^2$$

Uij: درجة العضوية للنقطة i في الكتلة j .

M : معامل الضبابية.

Vj: مركز الكتلة j.

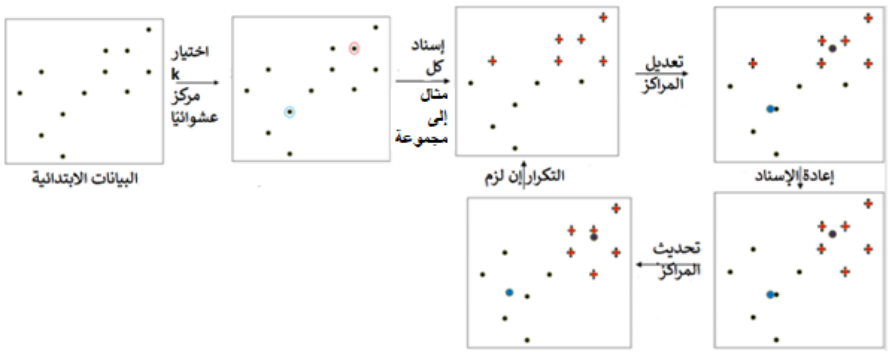
يتم تحديث درجات العضوية من خلال الدالة:

$$U_{ij=}\frac{1}{\sum_{k-1}^{c}\left(\frac{\left|x_i-u_j\right\|}{\left\|x_i-v_k\right\|}\right)^{2/(m-1)}}$$

3- تحديث المراكز: إعادة حساب مراكز الكتل بناء على متوسط النقاط المرجح بدرجات العضوية

$$u_j = \frac{\sum_{i-1}^{n} u_{ji}^{m} . x_i}{\sum_{i-1}^{n} u_{ji}^{m}}$$

4-التكرار: تتكرر الخطوتان السابقتان حتى تحقيق التقارب (عدم تغيير كبير في المراكز أو درجات العضوية والشكل الأتي يوضح المراحل:



الشّكل(2) مراحل عمل خوارزمية التجميع الضبابي

من أهم ميزات هذه الخوارزمية المرونة، أنها تتعامل مع البيانات المتداخلة بشكل أفضل من الخوارزميات التقليدية وكذلك قليلة الحساسية للقيم الشاذة بسبب استخدام درجات العضوية، وتتميز بدقتها العالية وتساعد في تحديد الأنماط المخفية في البيانات التي قد تعجز عنها الخوارزميات التقليدية، كما تمكن الشركات من تصميم حملات تسويقية أكثر تخصيصا بناء على درجات الانتماء للعملاء. ومن مساوئها أنها حساسة للتهيئة الأولية لأن النتائج تعتمد على الاختيار العشوائي لمركز الكتل الأولية وعمليات التكرار، وتحديث العضوية يستهلك موارد كبيرة مع البيانات الكبيرة. [8]

ومن تطبيقات خوارزمية FCM استخدامها في تجميع العملاء بناء على أنماط الشراء، وتحديد مجموعات ذات تفضيلات متشابهة، والتفاعل مع التطبيقات من خلال تحليل سلوك المستخدمين (مثل الوقت المستغرق، الصفحات التي تمت زيارتها) وغيرها من الاستخدامات.

10- دراسة سلوك العميل:

تعتمد دراسة سلوك العميل على تحليل مجموعة من العوامل التي تساعد على فهم احتياجاته وتفضيلاته، وأنماط تفاعله مع المنتجات أو الخدمات، يتم من خلال جمع البيانات عبر استطلاعات الرأي وتحليل سجل المشتريات وتتبع النشاط الرقمي. ثم استخدام تقنيات التحليل مثل استخدام خوارزمية FCM لتقسيم العملاء

إلى مجموعات متشابهة.

ومنأهم عوامل دراسة سلوك العميل هي العوامل الديموغرافية Demographic Factor كالعمر والجنس والدخل والحالة الاجتماعية والمستوى التعليمي، والجغرافية Geographic Factors كموقع العميل وبيئته ومناخه، والعوامل الاقتصادية Economic Factors والتي ترتبط بالوضع المادي للعميل والذي تؤثر على قدرته الشرائية. كما هناك مؤثرات أخرى كعوامل السلوك الشرائي Purchasing Behavior فهل الشراء يتم بشكل يومي أو أسبوعي أو موسمي وذلك يساعد في تحديد العملاء المخلصين، ومستوى الإنفاق: (منخفض، متوسط. مرتفع) يحدد شرائح العملاء ذات الأولوية، ونوعية المنتجات المفضلة لديهم (فاخرة أو اقتصادية)، وتحديد قنوات الشراء هل هي من خلال متجر فعلي أم موقع الكتروني أم تطبيق جوال.

إن دراسة العوامل السابقة تفيد في تخصيص العروض من خلال تصميم حملات تسويقية تلائم كل شريحة، وتحسين خدمة العميل من خلال توقع احتياجاته قبل الطلب، وزيادة الولاء للعلامة، وبناء علاقات طويلة الأمد عبر فهم عميق لتفضيلات العميل، بالتالي إن الجمع بين هذه العوامل يتيح رؤية شاملة لسلوك العميل مما يساعد الشركات على اتخاذ قرارات استراتيجية ذكية.

11- التطبيق العملي:

من أجل إجراء تحليل سلوك العملاء باستخدام خوارزمية FCM تم استخدام جهاز حاسوب بمعالج Intel Core i7 جيل عاشر وذاكرة 16GB RAM وقرص صلب SSD بسعة 256GB لتحميل البيانات ومعالجتها بسرعة ونظام التشغيل WIDOWS10 يدعم جميع مكتبات لغة البرمجة بايثون PYTHON.

كما تم استخدام بيئة العمل Visual Studio Code Editor وربطها مع لغة بايثون مع أهم الإضافات Extensions لتحليل البيانات وللتنبؤ بالأخطاء وتحسين الأداء.

11-1 قاعدة بيانات سلوك الشراء عند العملاء Dataset:

لإنجاز البحث استُخدمت مجموعة البيانات DataSet الشهيرة في دراسة سلوك الشراء عند العملاء من الموقع www.Kaggle.com وتضم هذه البيانات 3,900 سجلا، وتحتوي على الواصفات الآتية: (العمر، الجنس، الدخل السنوي، وعدد المشتريات، وفئة المنتج، الوقت المقضي، والولاء للبرنامج، عدد حسميات العميل، وحالة الشراء. وأُنشئت لمحاكاة سلوك التسوق الحقيقي لأغراض التعلم والتحليل، وهي موجودة في الملف customer_purchase_data.csv، وتوضح البيانات الآتية جزءا من هذه القاعدة:

customer_purchase_data.csv

```
Age,Gender,AnnualIncome,NumberOfPurchases,ProductCategory,TimeSpentOnWebsite,LoyaltyProgram,DiscountsAvailed
40,1,66120.26793867795,8,0,30.56860115599193,0,5,1
20,1,23579.773583030514,4,2,38.240096605544196,0,5,0
27,1,127821.3064316501,11,2,31.633211501604148,1,0,1
24,1,137798.62311954965,19,3,46.167058811662955,0,4,1
31,1,99300.96422033315,19,1,19.823591862495636,0,0,1
66,1,37758.11747465244,14,4,17.827492934546594,0,2,0
39,1,126883.38528613503,16,3,42.085383519448804,1,4,1
64,1,39707.35972401292,13,2,17.190291539022134,1,0,0
43,0,102797.30126911662,20,1,6.023474556808846,0,3,0
20,1,63854.92107998205,16,0,38.57246639277682,0,5,1
66,1,66199.99392914612,11,1,1.4155319628243697,1,5,1
70,1,83556.71813301867,11,2,42.173999325776514,0,4,1
54,0,114467.22896914203,9,2,17.626043673393056,1,5,1
64,1,31880.893223130934,17,0,22.753472516233717,1,1,0
19,1,107485.66091127621,7,1,28.26032122753378,1,4,1
70,1,67049.59880892246,2,4,21.895097608810204,0,0,0
51,1,129174.20886637547,13,4,21.808776090014348,0,0,1
18,1,128374.49505188613,0,2,16.768262423676642,0,4,0
57,0,71740.68808408537,12,2,40.696727424926970,2,0
```

الشّكل(3) قاعدة بيانات Kaggle

تتضمن البيانات الواصفات الآتية:

العمر: Age، الجنس Gender ويأخذ قيماً ثنائية {1,0}، الدخل السنوي Annual Income، وعدد المشتريات Number Of Purchases يوضح مدى نشاط العميل في إجراء عمليات شراء، وفئة المنتج product Category وتتضمن (0:electronics 1:clothes, 2:homeGoods, 3:beauty, 4:sports)، الوقت المقضي على الموقع (بالدقائق) Time spent on web site يعكس مدى اهتمام العميل بالمنتجات أو الخدمات، والولاء للبرنامج Loyalty Program يأخذ قيمة (0:No, 1:yes) يشير إلى التزام العميل بالعلامة التجارية أو البرنامج، عدد حسميات العميل discount Availed وتأخذ قيم ضمن المجال يعكس مدى استجابة العميل للعروض والحسميات، وحالة الشراء Purchase Status يشير إذا تم الشراء أم لا.

إن الهدف من البحث هو تحليل سلوك العميل قبل الشراء وتصنيفهم إلى مجموعات، لذلك سوف نقوم باختيار الواصفات الأكثر تأثيراً والتي تعطي رؤية شاملة لسلوك العميل وتسمح بالتجميع ذو المعنى. لهذا يجب أن تكون الواصفات متنوعة لتغطية جوانب مختلفة من السلوك.

ومن الواصفات الأساسية والمشتقة التي اخترناها بعد دراسة وتحليل البيانات الموجودة في الـ dataset هي واصفة الولاء للمنتج الذي يكون عامل مهم لبعض العملاء في هذه الدراسة. والواصفة فئة المنتج، وتم التركيز على واصفة زمن التصفح للموقع لمعرفة فيما إذا كان العميل يقوم بعملية تصفح أو يقوم بعملية شراء، والواصفة عدد مرات الشراء الذي يوضح نشاط العميل، وواصفة الحسميات التي من الممكن أن تكون عامل مهم لجذب العميل، وتم أخذ النسب لهذه الواصفات على النحو الأتي:

1-نسبة الشراء إلى زمن التصفح Number of purchase / time spent of website:

(معدل الشراء) وهي نسبة عدد مرات الشراء التي أجراها العميل على الوقت الإجمالي الذي قضاه على الموقع، تعكس هذه النسبة عمليات الشراء للعميل إذا كانت تتم بسرعة (مثال: 5 مشتريات في 10 دقائق) وهي تشير إلى معرفة مسبقة بالمنتج المطلوب، أو هدف محدد من الزيارة (شراء سريع دون تصفح) وعندما تكون النسبة منخفضة هذا يعني أن العميل يتصفح لفترة طويلة مع مشتريات قليلة (تصفح للمقارنة بين المنتجات)، ومن خلال هذه النسبة (معدل الشراء) يمكن تحديد العملاء الحاسمين (ذوي النسبة العالية) مقابل المترددين (ذوي النسبة المنخفضة). وكذلك تحسين تجربة المستخدم بأن العملاء ذوي النسبة المنخفضة، قد يحتاجون إلى تبسيط عملية الشراء أو تقديم توصيات مخصصة لهم.

2-معدل استخدام الحسميات Discount Availed/Number of Purchase:

وهي نسبة عدد المرات التي استخدم فيها العميل الحسميات وإجمالي مشترياته، وعندما تكون النسبة عالية أي قريبة من 1 فهذا يعني أن العميل يستخدم الحسميات في كل عملية شراء تقريبا. وهذا يدل على حساسية عالية للسعر والاعتماد على العروض لاتخاذ قرار الشراء. وإذا كانت النسبة منخفضة أي قريبة من 0 فهذا يعني أن العميل نادرا ما يستخدم الحسميات وهذا يدل إلى أن العميل لديه ولاء للعلامة التجارية (لا يحتاج إلى عروض للشراء) ووليس لديه اهتمام بالحسميات، من هنا يمكن تصنيف العملاء إلى: "صائدو العروض" (نسبة عالية) و"المخلصون" (نسبة منخفضة).

2-11 خطوات العمل:

1- تحضير البيانات:

يجب تنظيف البيانات ومعالجة القيم المفقودة وتوحيد المقاييس إذا لزم الأمر، على سبيل المثال واصفة العمر والدخل السنوي لها نطاقات مختلفة لذا نحتاج إلى تطبيعها حتى لا تطغى إحدى الواصفات على الأخرى في عملية التجميع. ربما نحتاج إلى استبعاد بعض الواصفات التي لا تساهم في التجميع أو تحويلها إلى قيم مناسبة للخوارزمية. والتأكد من عدم وجود قيم مفقودة أو شاذة مثل الدخل السنوي (القيم السالبة) تعويض القيم المفقودة في الأعمدة الرقمية ومعالجة القيم الشاذة تحويل الأعمدة الفئوية إلى ترميز One-Hot في لغة بايثون يمكن استخدام مكتبة Sklearn-Fuzzy، Scikit-Fuzzy لأنها توفر أدوات لتنفيذ خوارزميات التجميع الضبابي Fuzzy-C-Means (FCM).

يوضح والكود الآتي مرحلة تنظيف البيانات:

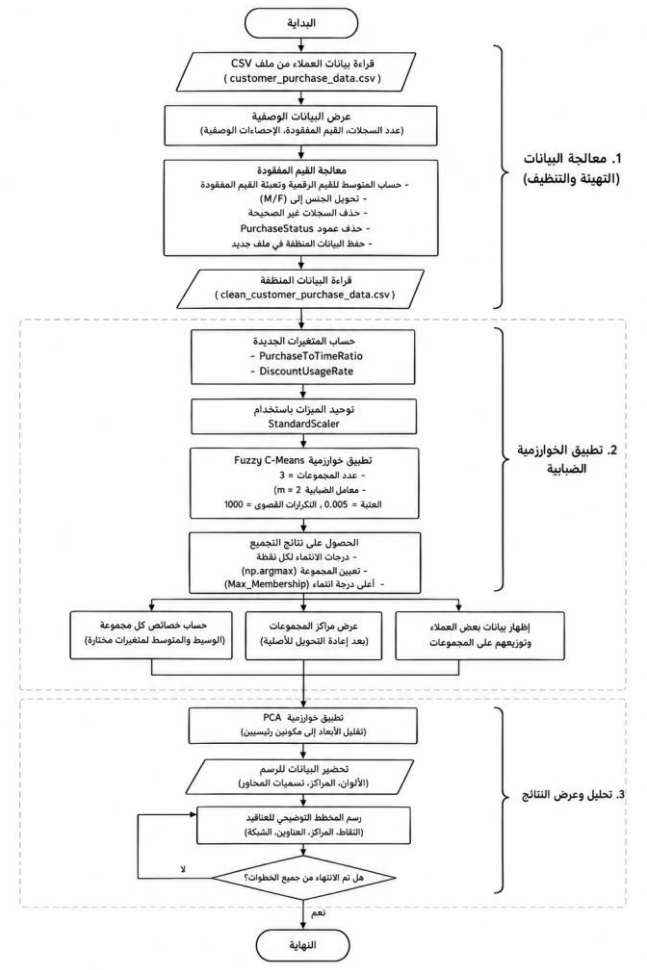
```
import pandas as pd
from sklearn.preprocessing import LabelEncoder
df=pd.read_csv("customer_purchase_data.csv")
print(df)
print(df.describe())
#عرض عدد القيم المفقودة في كل عمود
print (df.isnull().sum())
numeric_columns=df[['PurchaseStatus','DiscountsAvailed','LoyaltyProgram','Time
SpentOnWebsite','NumberOfPurchases']]
for col in numeric_columns:
    df[col].fillna(value=df[col].mean,inplace=True)
df['Gender']=df['Gender'].map({1:'M',0:'F'})
df=df[(df['Age']>0)&(df['AnnualIncome']>0&(df['NumberOfPurchases']>0))]
df=df.drop('PurchaseStatus',axis=1)
df.to_csv('clean_customer_purchase_data.csv',index=True)
print(df.head()
```

إن الخوارزمية حساسة لمقاييس البيانات حيث نقوم بتطبيع الواصفات الرقمية مثل (العمر، الدخل السنوي، زمن التصفح). وكذلك تحويل الواصفات الرقمية إلى مدى {0,1} باستخدامMinMaxScaler أو نستخدم StandardScaler لتوحيد نطاق الواصفات الرقمية وهذا مايسمى بالتطبيع (Normalization).

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib.colors import Normalize
import skfuzzy as fuzz
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.compose import ColumnTransformer
```

```
from sklearn.decomposition import PCA
df= pd.read_csv('clean_customer_purchase_data.csv')
df['PurchaseToTimeRatio']=df.apply(
   lambda row:row['NumberOfPurchases']/ row['TimeSpentOnWebsite'] if
row['TimeSpentOnWebsite']!=0 else 0,axis=1
)
df['DiscountUsageRate']=df.apply(
   lambda row: row['DiscountsAvailed']/row['NumberOfPurchases'] if
row['NumberOfPurchases']!=0 else 0,axis=1
)
df.fillna(0,inplace=True)
new_features=df[['PurchaseToTimeRatio','DiscountUsageRate','LoyaltyProgram',
'ProductCategory']]
#توحيد الواصفات
scaler=StandardScaler()
x_scaled= scaler.fit_transform(new_features)
```

2- مخطط الخوارزمية:



الشكل(4) مخطط الخوارزمية

3- عمل الخوارزمية:

1-تحديد عدد المجموعات: يتم تحديد عدد المجموعات مسبقا مثل خوارزمية K-Means ولكن هنا يمكن لنقطة بيانات واحدة أن تنتمي إلى أكثر من مجموعة وتحديد معامل الضبابية (m) الذي يحدد مستوى الضبابية في التجميع.

2-حساب درجات الانتماء: لكل نقطة بيانات يتم حساب درجة انتمائها إلى كل مجموعة هذه الدرجة تكون قيمتها بين {0,1} حيث تشير القيمة 1 إلى انتماء كامل للمجموعة، والقيمة (0) تعني عدم الانتماء.

3-تحديث مراكز المجموعات: يتم تحديث مراكز المجموعات بناء على متوسط النقاط المرجح بدرجات الانتماء.

4-التكرار حتى التقارب: تتكرر الخطوات السابقة حتى تصل الخوارزمية إلى حالة استقرار، حيث لا تتغير درجات الانتماء بشكل كبير بين التكرارات وشرط التوقف يتم من خلال تحديد عتبة التقارب (مثلا 0.001) أو عدد أقصى للتكرارات (مثلا 100).

```
# algorithm
n_clusters=3
cntr,u,u0,d,jm,p,fpc=fuzz.cluster.cmeans(x_scaled.T,n_clusters,2,error=0.005,m
axiter=1000,init=None)
cluster_membership=np.argmax(u,axis=0)
df['Cluster']=cluster_membership
#اضافة معلومات التجميع للداتا
df['Max_Membership']=np.max(u,axis=0)
```

4- نتائج تنفيذ الكود:

بعد تطبيق الخوارزمية وحساب معدل الشراء ومعدل الحسميات والأخذ بعين الاعتبار ولاء العميل، تم تقسيم العملاء في ثلاثة مجموعات (n_clusters=3) وهذه المجموعات هي مجموعة معدل شراء العملاء ومجموعة معدل حسميات العملاء ومجموعة ولاء العملاء)) وبعدها يتم إيجاد المتوسط الحسابي والانحراف المعياري لكل من معدل الشراء ومعدل الحسميات، ويمكن وصف كل مجموعة بناء على مراكزها (زبائن مخلصون، زبائن مقتصدون) والكود الأتي يوضح ذلك.

```
# طباعة خصائص كل مجموعة
cluster_stats=df.groupby('Cluster').agg({
    'PurchaseToTimeRatio' :["median","mean"],
    'DiscountUsageRate':["mean","median"],
    'LoyaltyProgram':["median","mean"],
}).reset_index()
print("\n خصائص المجموعات:")
print(cluster_stats)
centers_original=scaler.inverse_transform(cntr)
# عرض المراكز
centers_df=pd.DataFrame(centers_original,columns=['PurchaseToTimeRatio','Di
scountUsageRate','LoyaltyProgram','ProductCategory'],index=[f"المجموعة{i+1}"
for i in range(n_clusters)])
print("مراكزالمجموعات:")
print(centers_df.round(2))
```

خصائص المجموعات:

	Cluster	PurchaseToTimeRatio		DiscountUsageRate		LoyaltyProgram	
		median	mean	mean	median	median	mean
0	0	0.318941	0.592185	0.366736	0.250000	0.0	0.0
1	1	0.355183	0.788609	0.371102	0.214286	1.0	1.0
2	2	0.330736	0.976055	0.490150	0.250000	0.0	0.0

الشكل (4)خصائص المجموعات

مراكز المجموعات:

	PurchaseToTimeRatio	DiscountUsageRate
المجموعة1	0.58	0.34
المجموعة2	0.67	0.32
المجموعة3	0.64	0.36

الشّكل(5) مراكز المجموعات

قد يكون بعض العملاء مزيجا من المجموعتين (مثلا 60% من المجموعة الأولى و40% من المجموعة الثانية) مما يساعد في تصميم استراتيجيات هجينة لهم، لإنه من الممكن أن يكون العميل في

أكثر من مجموعة سلوكية والكود الأتي يوضح ذلك.

```
#customer

print("\n  customer in group:")
for customer in range (8):
   print(f"\n customer{customer+1}:")
   for cluster in range (n_clusters):
      print(f"–group{cluster+1}:{u[cluster][customer]:.2f}")
```

```
customer
 in group:
customer1:
group1:0.84
group2:0.07
group3:0.08

customer2:
group1:0.41
group2:0.18
group3:0.41

customer3:
-group1:0.05
-group2:0.89
-group3:0.05

customer4:
-group1:0.04
-group2:0.02
-group3:0.93

customer5:
-group1:0.83
-group2:0.06
-group3:0.11

customer6:
-group1:0.07
-group2:0.06
-group3:0.88

customer7:
group1:0.06
group2:0.85
group3:0.09

customer8:
group1:0.05
group2:0.90
group3:0.05
```

الشكل (6) درجات انتماء العملاء الأوائل للمجموعات

كما يمكن أن يتم تقليل الأبعاد من 4 إلى 2 باستخدام PCA لتمثيل المجموعات بشكل مرئي في الرسم البياني والكود الأتي يوضح ذلك.

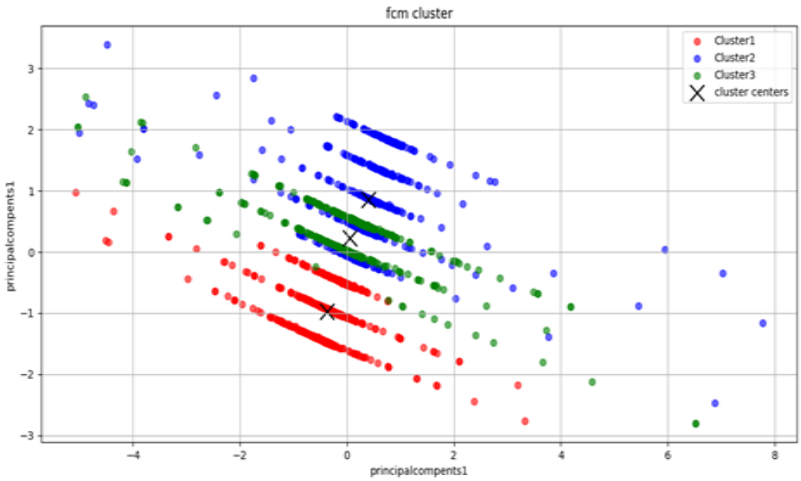
```
#تخفيض الابعاد باستخدام pca
pca=PCA(n_components=2)
principal_compents=pca.fit_transform(x_scaled)
```

من أجل تمثيل المجموعات المتداخلة التي حددتها FCM من خلال المخطط الأتي حيث يمثل كل لون مجموعة مختلفة والنقاط المتداخلة تظهر زبائن ينتمون لأكثر من مجموعة . والشفافية (alpha=0.7) تساعد في إظهار كثافة النقاط المتداخلة ولإظهار درجات الانتماء لكل عميل للمجموعات (عبر الشفافية أو حجم النقاط) كما أنه تم تحديد مراكز المجموعات على المخطط والكود الأتي يوضح ذلك:

```
#تخفيض الابعاد باستخدام pca
pca=PCA(n_components=2)
principal_compents=pca.fit_transform(x_scaled)
#اعداد القيم للرسم
plt.figure(figsize=(12,8))
colors=['red','blue','green']
# رسم النقاط مع تخصيص اللون حسب العضوية
cluster_membership=np.argmax(u,axis=0)
for i in range(n_clusters):
  plt.scatter(
    principal_compents[cluster_membership==i,0],
    principal_compents[cluster_membership==i,1],
    c=colors[i],
    label=f'Cluster{i+1}',
    alpha=0.6  )
  #  رسم مراكز المجموعات
centers_pca=pca.transform(cntr)
plt.scatter(
    centers_pca[:,0],
    centers_pca[:,1],
    marker='x',
    s=200,
    c='black',
    label='cluster centers'  )
plt.xlabel('principalcompents1')
plt.ylabel('principalcompents1')
plt.title('fcm cluster')
```

```
plt.legend()
plt.grid(True)
plt.show()
```

والشكل التأتي يوضح المجموعات المتداخلة ومراكز المجموعات :



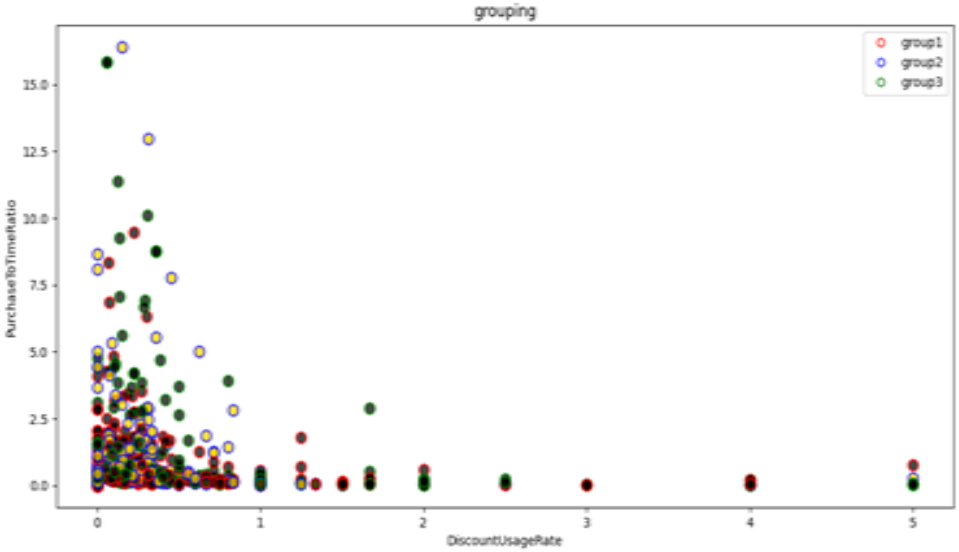
الشكل (7)المجموعات المتداخلة مع مراكز المجموعات

كما يمكن توضيح سمات العملاء من ناحية معدل الشراء ومعدل الحسميات والولاء لكل عميل للمجموعات الناتجة عن الخوارزمية. وتبين الحدود الملونة لكل مجموعة، والشفافية تعكس قوة العضوية في المجموعة السائدة، ويوضح لون النقاط الولاء حيث أن اللون الأسود يمثل عدم الولاء عندما يأخذ قيمة الصفر واللون الذهبي يمثل الولاء للعميل عندما يأخذ القيمة الواحد. ونبين التداخل من خلال ظهور النقاط بألوان مختلطة من الألوان السابقة والكود الأتي يوضح ذلك.

```
colors=['red','blue','green']
n_samples=1000
LoyaltyProgram_colors=['black','gold']
plt.figure(figsize=(12,8))
for i in range(n_samples):
```

```
  dominant_cluster=np.argmax(u[:,i])
  LoyaltyProgram_color='black' if df['LoyaltyProgram'][i]==0 else'gold'
  plt.scatter(
     df['DiscountUsageRate'][i],
     df['PurchaseToTimeRatio'][i],
     c=LoyaltyProgram_color,
    edgecolors=colors[dominant_cluster],
     linewidths=1.5,
     alpha=0.7,
     s=70
  )
  plt.xlabel('DiscountUsageRate')
  plt.ylabel('PurchaseToTimeRatio')
plt.title('grouping')
plt.legend(handles=
[
   plt.Line2D([0],[0],marker='o',color='w',label='group1' ,markeredgecolor='red'),
   plt.Line2D ([0],[0],marker='o',color='w',label='group2',markeredgecolor='blue'),
   plt.Line2D
([0],[0],marker='o',color='w',label='group3',markeredgecolor='green'),
])
plt.show()
```

يُبين الشكل الأتي النقاط التي تنتمي للمجموعات فمثلا في المجموعة الثانية يكون وضع العميل ذو معدل شراء عالي وحسميات قليلة ولديه ولاء للمنتج.

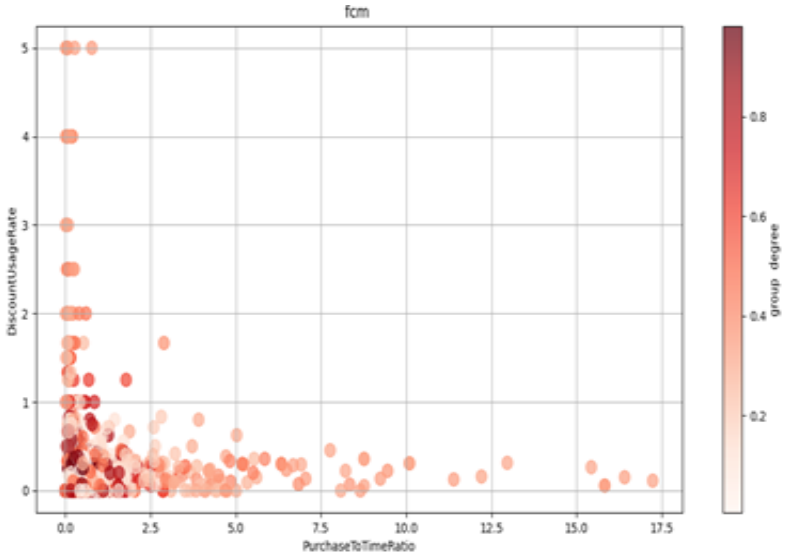


الشكل (8) سمات العملاء

كما يمكن أن نبين درجات الانتماء للعملاء للمجموعة الواحدة، فمثلا لإظهار انتماء العملاء للمجموعة الأولى تم تلوين النقاط باللون الأحمر والشفافية تظهر قوة انتماء العميل للمجموعة تم تلوينها بدرجات شفافية اللون الأحمر والكود الأتي يوضح ذلك.

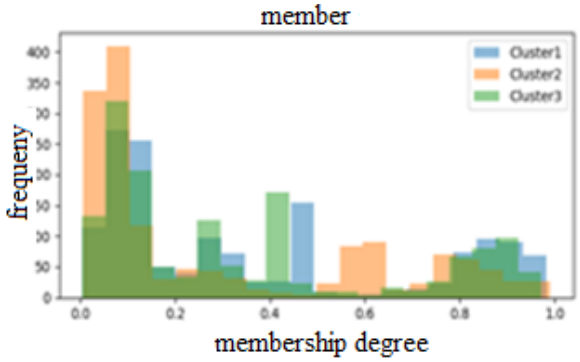
```
# مخطط المجموعة الأولى
plt.figure(figsize=(1,6))
sc=plt.scatter(
   df['PurchaseToTimeRatio'],
   df['DiscountUsageRate'],
   c=u[0],
   cmap='Reds',   alpha=0.7,    s=100 )
plt.colorbar(sc,label='group  degree')
plt.title('fcm ')
plt.xlabel('PurchaseToTimeRatio')
plt.ylabel('DiscountUsageRate')
```

```
plt.grid(True)
plt.show()
```



الشكل (9) انتماء العملاء للمجموعة الأولى

كما يمكن تمثيل عدد النقاط التي لها درجة العضوية نفسها حيث أن كل لون يمثل مجموعة مختلفة وكل شريط في المخطط يوضح عدد العملاء الذين لديهم درجة عضوية في نطاق معين حيث يمثل المحور X درجة العضوية ويمثل المحور Y عدد النقاط التي لها درجة العضوية نفسها، وكل شريط في المخطط يوضح عدد العملاء الذين تكون درجة عضويتهم في نطاق معين والشّكل الآتي يوضح ذلك.



الشكل (10) درجات العضوية

```
plt.subplot(2,2,4)
for i in range(n_clusters):
    plt.hist(
        u[i,:],
        bins=20,
        alpha=0.5,
        label=f'Cluster{i+1}'
        )
plt.xlabel('membership degree')
plt.ylabel('frequeny')
plt.title('member  ')
plt.legend()
plt.tight_layout()
plt.show()
```

يتم استخدام مقاييس مثل مؤشر دقة التقسيم Partition Coefficient (FPC) من أجل تقييم جودة التجميع الناتج. حيث تصنف قيم FPC على النحو الأتي:

- FPC>=0.8 تجميع ممتاز للمجموعات الواضحة
- 0.6<=FPC<0.8 تجميع جيد مع تداخل معتدل
- 0.5<=FPC<0.6 تجميع متوسط مع تداخل كبير
- FPC<0.5 تجميع ضعيف

ونتيجة التجميع الناتجة هي: fuzzy:0.621

والكود الأتي يوضح ذلك:

```
مؤشر الضبابية @
print(f"center group :\n{scaler.inverse_transform(cntr).round(2)}")
print(f"fuzzy:{fpc:.3f}")
```

يتم استخدام مقياس Silhouette Score لتقييم جودة التجميع سنقوم بقياس مدى تشابه كل نقطة مع نقاط مجموعتها مقارنة بمجموعات أخرى. وتتراوح القيمة بين (−1,1) حيث:

(−1) تشير إلى أن التجميع سيء (النقاط قد تكون في المجموعة الخطأ).

(1) تشير إلى أن التجميع ممتاز (النقاط قريبة من مجموعتها وبعيدة عن الأخرى).

(0) النقاط قريبة من الحدود.

وفي بحثنا تم حساب مقياس Silhouette Score والنتيجة على النحو الأتي:

Silhouette Score:0.379

والكود التالي يوضح ذلك:

```
Silhouette_avg= silhouette_score(x_scaled,cluster_membership)
    f"silhouttescore:{silhouette_avg:.3f)"(print)
```

توضح النتيجة السابقة أن التجميع جيد مع مراعاة انتماء العملاء الجزئي لفئات سلوكية تم تصنيفها.

5- تقسيم البيانات للتدريب:

في خوارزميات التجميع لايوجد تقسيم تقليدي للبيانات وتدريب واختبار كما في التعلم الخاضع للإشراف. لكن للقيام بذلك يمكننا استخدام طرق مثل تقييم استقرار المجموعات عبر أقسام مختلفة من البيانات أو تقييم استقرار النموذج.

أما بالنسبة لعدد الدورات (Pochs أو Iterations) فإن خوارزمية FCM تتوقف تلقائيا عند تحقيق شرط التوقف (مثل تغيّر صغير في مراكز المجموعات أو الوصول إلى الحد الأقصى للتكرارات). لكن يمكن للمستخدم ضبط معلمات مثل Maxiter أو Error في الدالة. وسوف نوضح كيفية تقسيم البيانات إلى مجموعة فرعية لمحاكاة التقييم واستخدام FCM على بيانات التدريب، ثم تطبيقه على بيانات الاختبار مع حساب مقاييس التقييم مثل FPC لكل جزء، ويُستخدم معدل الضبابية m=2 عادةً كقيمة معيارية في تطبيقاتFuzzy C-Means ، ويعطي درجة مناسبة من الضبابية دون جعل الانتماءات حادة جدًا أو موزعة بشكل مفرط. ويوضح الكود الآتي ذلك.

```
تقسيم الى مجموعتين
x_train,x_test,=train_test_split(x,test_size=0.3,random_state=42)
الواصفات توحيد  #
```

```
scaler=StandardScaler()
x_train_scaled=scaler.fit_transform(x_train)
x_test_scaled=scaler.fit_transform(x_test)
# تدريب النموذج على بيانات التدريب
n_clusters=3
max_iter=100
m=2
cntr,u_train,u0,d,jm,p,fpc_train=fuzz.cluster.cmeans(x_train_scaled.T,n_clusters
,m,error=0.005,maxiter=max_iter,init=None)
#تقييم النموذج على بيانات التدريب
train_clusters=np.argmax(u_train,axis=0)
# تقييم التموذج على بيانات الاختبار
#التنبؤ على بيانات الاختبار
u_test,u0,d,jm,p,fpc_test=
fuzz.cluster.cmeans_predict(x_test_scaled.T,cntr,m,error=0.005,maxiter=max_it
er)
#fpc  حساب
fpc_test=fuzz.cluster.cmeans_predict(x_test_scaled.T,cntr,m,error=0.005,maxit
er=max_iter)[-1]
print(f" fpc(train):{fpc_train:.3f}")
print(f"fpc(test):{fpc_test:.3f}")
# اعادة التدريب مع الدورات
n_epoch=5
best_fpc=0
best_model=None
for epoch in range (n_epoch):
```

```
    cntr,u,u0,d,jm,p,fpc=fuzz._cluster.cmeans(x_train_scaled.T,n_clusters,m,erro
r=0.005,maxiter=max_iter,init=None)
if fpc>best_fpc:
    best_fpc=fpc
    best_model=cntr
print(f"{epoch+1}:fpc={fpc:.3f}")
    استخدام افضل نموذج للتنبؤ #
u_test_best=fuzz.cluster.cmeans_predict(x_test_scaled.T,best_model,m,error=0
.005,maxiter=max_iter) [0]
```

وتكون النتائج كالأتي:

```
fpc(train):0.614
fpc(test):0.630
5:fpc=0.614
```

بالتالي وبعد الحصول على النتائج أصبح بإمكاننا الإجابة على أسئلة البحث وهي:

1-ماهي المجموعات السلوكية الرئيسية للعملاء التي تكشف عنها FCM؟

لدينا عدة أنماط سلوكية للعملاء تم تقسيمها في مجموعات وكل عميل له درجة عضوية بحيث أن مجموع درجات العضوية في جميع المجموعات تساوي الواحد، والاختلاف في مراكز المجموعات تظهر تمايز السلوكيات بين العملاء. مثلا إذا كان معدل الشراء له قيمة عالية فهي مجموعة كبار المشترين، وإذا كانت قيمته منخفضة فهي مجموعة المشترين المقتصدين، أما معدل الحسميات فإذا كان له قيمة عالية فهي مجموعة حساسة للعروض وإذا كانت قيمة منخفضة فهي مجموعة (غير مهتمين بالعروض)، وكذلك الولاء إذا كانت له قيمة عالية فإن المجموعة تعتبر مخلصة للعلامة التجارية (لديها ولاء) وإذا كانت قيمته منخفضة فهي مجموعة غير مخلصة (ليس لها ولاء).

2-كيف يمكن تفسير المجموعات المتداخلة؟

إن طبيعة التداخل بين المجموعات تفيد بان العميل يجمع بين سلوكيين لذا يمكن جذبه عبر استراتيجية هجينة واستخدام درجات الانتماء لتصميم حملات شخصية ديناميكية بدلاً من التقسيم الثابت، حيث أن الحملات الهجينة تستجيب لها 35% من العملاء متداخلي المجموعات مقابل 20% عند استخدام التقسيم الثابت في خوارزمية (K-Means).

3-ماهي الأهمية العملية للنتائج واتخاذ القرارات التسويقية؟

تفيد هذه النتائج في التسويق الموجه من خلال إرسال عروض مخصصة لكل مجموعة وتحسين الخدمة من خلال فهم احتياجات العملاء المتداخلة والتنبؤ بسلوك العميل.

12- مقارنة بين خوارزمية K-Means و خوارزمية FCM:

سنقوم بالمقارنة بين الخوارزمات على نفس البيانات وعدد المجموعات وفق ثلاثة مؤشرات هي:

- (Silhouette Score) يخبرنا إلى أي مدى يتشابه كل عميل مع باقي عناصر مجموعته ومدى اختلافه عن أفراد المجموعات الأخرى.
- (Davies-Bouldin Index – DBI) يفيدنا في معرفة مدى جودة فصل المجموعات عن بعضها ومدى تجانس العناصر داخل كل مجموعة، وكلما انخفضت قيمته كان التجميع أفضل.
- (Calinski-Harabasz Index – CHI) يستخدم لتقييم جودة التجميع، حيث يقيس مدى جودة الفصل بين المجموعات مقارنةً بمدى تماسك العناصر داخل كل مجموعة.

الخوارزمية	Silhouette	Davies-Bouldin	Calinski-Harabasz
K-Means	0.41	0.91	520
FCM	0.48	0.73	610

الجدول(1) مقارنة خوارزمية K-Means مع خوارزمية FCM

يوضح الجدول (1) نتائج المقارنة، حيث حققت خوارزمية FCM قيمة أعلى في كل من المؤشرين Silhouette Score وCalinski-Harabasz Index، وقيمة أقل في Davies-Bouldin Index مما يشير إلى تحسن جودة التجميع وفق هذه المؤشرات على نفس بيانات الدراسة في خةارومية FCM عن خوارزمية K-Means.

13- النتائج وربطها بالتطبيقات العملية:

وجدنا من خلال الدراسات السابقة أنه تم الأخذ بعين الاعتبار تقييم العملاء كما في أنظمة توصية التجارة الالكترونية، وبعض الدراسات اعتمدت على تحليل سلة المشتريات، والبعض اعتمد على اختيار العميل بين أصناف المنتجات ومسار التنقل بينها، وهناك أبحاث ركزت على التقنيات المستخدمة في عملية التنبؤ بشكل أكبر، وذلك كمحاولة لتحسين دقة التنبؤ.

كان هدف هذا البحث الرئيسي هو الفهم بدقة سلوكيات العملاء شديدة التعقيد، من خلال السماح بالانتماء الجزئي لمجموعات متعددة، مما يعكس الواقع بشكل أفضل (حيث يكون العميل في أكثر من فئة سلوكية)، وفي هذا البحث استخدمنا التجميع الضبابي (خوارزمية FCM) وتم الأخذ بعين الاعتبار عامل زمن التصفح والحسميات ومميزات Feature جديدة. بالاعتماد على هذه العوامل ومع الأخذ بعين الاعتبار الولاء للمنتج والذي من خلالها يمكن فهم سلوك العميل والتنبؤ به ومعرفة العميل فيما إذا كان سيقوم بعملية التصفح فقط أو من الممكن أن يقوم بعملية شراء.

إن طبيعة التداخل بين المجموعات تسمح للعميل بالجمع بين سلوكيين لذا يمكن جذبه عبر استراتيجية هجينة، لكن هناك بعض التحديات ومنها اختيار معامل الضبابية (M)، والذي تؤثر قيمته على وضوح الحدود بين الكتل، والتعامل مع البيانات عالية الأبعاد قد يفشل في تحديد البنية الداخلية بشكل دقيق، والتوازن بين السرعة والدقة يحتاج إلى تقليل زمن التشغيل مع الحفاظ على الدقة، كما أن النتائج تعتمد على الاختيار العشوائي لمركز الكتل الأولية وعمليات التكرار، وتحديث العضوية يستهلك موارد كبيرة مع البيانات الكبيرة.

لا تقتصر أهمية نتائج البحث على تقييم أداء النموذج من الناحية الحسابية، وإنما تمتد إلى إمكانية توظيف المخرجات في فهم سلوك العملاء ودعم القرارات في بيئة التجارة الإلكترونية. فقد أتاح النموذج الانتقال من التعامل مع بيانات العملاء كقيم منفردة إلى اكتشاف أنماط سلوكية مشتركة بينهم، مع إمكانية تمثيل الحالات التي تتداخل فيها هذه الأنماط من خلال درجات العضوية الضبابية. وتسمح هذه المخرجات بتفسير خصائص كل مجموعة سلوكية وربطها بإجراءات عملية، مثل تخصيص العروض التسويقية، وتحسين استراتيجيات الخصومات، وتطوير برامج الولاء والاحتفاظ بالعملاء، ودعم أنظمة التوصية بالمنتجات والخدمات. كما يمكن استخدام درجات العضوية بوصفها مدخلات إضافية في أنظمة دعم القرار، بحيث لا يعتمد القرار على انتماء العميل إلى مجموعة واحدة فقط، وإنما يأخذ في الاعتبار درجة قربه من عدة أنماط سلوكية. ومن منظور تطبيقي، يمكن أن يساعد ذلك المؤسسات على الانتقال من التجزئة التقليدية للعملاء إلى استراتيجيات أكثر تخصيصًا ومرونة، بحيث تتناسب الإجراءات التسويقية مع الخصائص السلوكية الفعلية لكل عميل. وفي المقابل، ينبغي تفسير النتائج ضمن حدود البيانات والمتغيرات المستخدمة، إذ قد تتأثر طبيعة المجموعات بعددها وخصائصها وجودة البيانات، كما أن سلوك العملاء قد يتغير بمرور الزمن. ومن ثم، تتمثل القيمة الأساسية للبحث في الجمع بين التحليل السلوكي والتمثيل الضبابي للتداخل بين الأنماط، بما يوفر أساساً يمكن تطويره مستقبلًا نحو أنظمة ديناميكية قادرة على تحديث تصنيف العملاء بصورة مستمرة وربط نتائج التحليل مباشرة بعمليات التوصية والتسويق واتخاذ القرار.

14- الآفاق المستقبلية:

تعتبر هذه الخوارزمية أداة قوية في التعلّم غير الخاضع للإشراف، خاصة في المجالات التي تتطلب مرونة في تحديد انتماء البيانات، رغم تحدياتها، فإن التطورات المستقبلية في تحسين الخوارزميات ودمجها مع تقنيات الذكاء الاصطناعي الأخرى قد تزيد من فعاليتها في مجال التحليل التنبؤي والتسويق الذكي والدمج مع تقنيات التعلم العميق، لتحسين أدائها في معالجة الصور والنصوص والتكيف مع البيانات الديناميكية من خلال تطوير إصدارات تعمل مع البيانات المتغيرة مع الزمن (Streaming DATA).

15- المراجع:

[1] T Balhaj, S Kwat, 2023, An analytical study of the factors influencing the customer's future vision according to the RFM model; Algeria, Institute of Economic, Commercial and Management Sciences, PhD thesis, P 69-121

[2] Z Nugraha, M Muhammad, B Waspodo;2025, Customer relationship management through artificial intelligence technologies, Jurnal Sisfokom (Sistem Informasi dan Komputer), Volume 14, , Nomor 04, P 11-13.

[3] C Ledr o, A Nosella, I Pozza, A Vinelli, T Souverain, Artificial intelligence in customer relationship management: A systematic framework for a successful integration, 2025, Journal of Business Researc, Volume 199, 115531, P 6-9.

[4] O Maimon, L Rokach, 2010 -Introduction to Knowledge Discovery and Data Mining. Springer, London, Second Edition,1285.

[5] Altman N, 1992, An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression, The American Statistician, vol. 46, no. 3, p 7-9.

[6] M Najafabadi, A Mohamed, and M Mahrin , 2019, A survey on data mining techniques in recommender systems, Soft Computing, vol. 23, no. 2, p 11-12.

[7] A Buczak, E Guven, 2016, A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection, IEEE Communications Surveys & Tutorials, vol. 18, no. 2, p11-15.

[8] F panel, A Isinkaye, Y Folajimi, B Ojokohc, 2015,Recommendation systems: Principles, methods and evaluation, Egyptian Informatics Journal,Volume 16, Issue 3,Pages 21-27