Libyan Open University Journal of Applied Sciences (LOUJAS)
Volume 1 - Issue 2 - 2025 - Pages 41-52

Applications of matrices in big data analysis: Modeling user behavior in digital systems

Eman Atia Ramadan Ali [1*], Naser Rafalla Naser [2], Nadya Abdallah Atlouba [3]
[1] Mathematics Department, Faculty of Education, Tobruk University, Libya
[2,3] Mathematics Department, Faculty of Science, University of Derna, Libya
***Email (for reference researcher):** bt20112016@gmail.com

تطبيقات المصفوفات في تحليل البيانات الضخمة: نمذجة سلوك المستخدمين في الأنظمة الرقمية

إيمان عطية رمضان علي [1*]، نصر راف الله نصر [2]، نادية عبد الله أطلوبة [3]
[1] قسم الرياضيات، كلية التربية، جامعة طبرق، ليبيا
[2،3] قسم الرياضيات، كلية العلوم، جامعة درنة، ليبيا

Received: 16-10-2025; Accepted: 05-12-2025; Published: 23-12-2025

Abstract:
This applied study aims to implement matrix-based techniques on real-world data to model user behavior in digital systems. The study was conducted on large-scale data extracted from an e-commerce platform comprising more than 100,000 users and 50,000 products. Non-negative Matrix Factorization (NMF) and similarity matrices were employed to analyze browsing and purchasing patterns. A predictive model based on principal component analysis of matrices was developed to forecast users' future behavior. The applied results demonstrated a predictive accuracy of 87.3% in anticipating user preferences and 82.1% in forecasting purchasing behavior. Furthermore, a recommendation system based on matrix analysis was developed, achieving a 34% improvement in click-through rate compared to traditional systems. This study provides a practical and applicable framework for leveraging matrix techniques in big data analysis for organizations and enterprises.

Keywords: Big Data, Matrices, User Behavior, Applied Study, Recommendation Systems.

الملخص
تهدف هذه الدراسة التطبيقية إلى تطبيق تقنيات المصفوفات الرياضية على بيانات حقيقية لنمذجة سلوك المستخدمين في الأنظمة الرقمية. تم تطبيق الدراسة على بيانات ضخمة مستخرجة من منصة تجارة إلكترونية تضم أكثر من 100,000 مستخدم و50,000 منتج، حيث تم استخدام تقنيات التحليل العاملي الكامن غير السالب (Non-negative Matrix Factorization) ومصفوفات التشابه لتحليل أنماط التصفح والشراء. تم تطوير نموذج تنبؤي يعتمد على تحليل المكونات الرئيسية للمصفوفات لتوقع سلوك المستخدمين المستقبلي. النتائج التطبيقية أظهرت دقة تنبؤية تصل إلى 87.3% في توقع تفضيلات المستخدمين، و82.1% في التنبؤ بسلوك الشراء. كما تم تطوير نظام توصيات يعتمد على تحليل المصفوفات حقق تحسناً بنسبة 34% في معدل النقر مقارنة بالأنظمة التقليدية. الدراسة تقدم إطاراً عملياً قابلاً للتطبيق لاستخدام المصفوفات في تحليل البيانات الضخمة للمؤسسات والشركات.

الكلمات المفتاحية: البيانات الضخمة، المصفوفات، سلوك المستخدمين، التطبيق العملي، نظم التوصيات.

- **المقدمة**

تواجه المؤسسات والشركات في العصر الرقمي تحدياً كبيراً في فهم وتحليل سلوك المستخدمين من خلال البيانات الضخمة التي تنتجها أنظمتها الرقمية يومياً. مع تزايد استخدام منصات التجارة الإلكترونية ومواقع التواصل الاجتماعي والتطبيقات المختلفة، تتراكم كميات هائلة من البيانات حول تفاعلات المستخدمين وسلوكهم (محمد، 2022). هذه البيانات تحتوي على معلومات قيمة يمكن أن تساهم في تحسين تجربة المستخدم وزيادة الأرباح، لكن استخراجها وتحليلها يتطلب استخدام تقنيات متقدمة وفعالة.

البيانات الضخمة في البيئات الرقمية تتميز بخصائص معقدة تجعل من التحليل التقليدي غير كافٍ لاستخراج الأنماط والاتجاهات المفيدة. فهي تتسم بالحجم الهائل حيث قد تحتوي قواعد البيانات على ملايين السجلات، والسرعة العالية في التحديث حيث تضاف آلاف التفاعلات كل دقيقة، والتنوع الكبير في أنواع البيانات من نصوص وصور وقيم رقمية وزمنية

(LOUJAS) **Copyright**: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

(يسري، 2023). كما تتميز البيانات بكونها متناثرة حيث لا يتفاعل كل مستخدم مع كل منتج، مما يخلق مصفوفات كبيرة بها نسبة عالية من القيم المفقودة أو الصفرية.

من الناحية التطبيقية، يعتبر تحليل البيانات الضخمة ضرورة حيوية للشركات الحديثة لفهم عملائها وتحسين خدماتها. فعلى سبيل المثال، تحتاج متاجر التجارة الإلكترونية إلى فهم تفضيلات المستخدمين لتقديم توصيات دقيقة، بينما تحتاج المنصات الاجتماعية إلى تحليل أنماط التفاعل لتحسين خوارزميات عرض المحتوى (أحمد، 2023). هذا الفهم العميق للسلوك يمكن الشركات من اتخاذ قرارات أكثر ذكاءً وتحقيق ميزة تنافسية في السوق.

تلعب المصفوفات الرياضية دوراً محورياً في تحليل هذه البيانات المعقدة، حيث توفر إطاراً رياضياً قوياً لتمثيل العلاقات متعددة الأبعاد بين المستخدمين والمنتجات والتفاعلات المختلفة. يمكن استخدام تقنيات المصفوفات لحل مشاكل عملية مثل التنبؤ بالتفضيلات، وتجميع المستخدمين ذوي السلوك المتشابه، واستخراج الأنماط المخفية في البيانات (جاسم، 2012). هذه التقنيات لا تقتصر على الجانب النظري فحسب، بل لها تطبيقات عملية مباشرة تؤثر على قرارات الأعمال والاستراتيجيات التسويقية.

- **مشكلة الدراسة وأهدافها**

تكمن المشكلة الأساسية في هذه الدراسة في الحاجة الماسة للمؤسسات والشركات لفهم سلوك مستخدميها من خلال تحليل البيانات الضخمة التي تجمعها أنظمتها الرقمية. تشير الدراسات إلى أن البيانات الضخمة تنمو بمعدل هائل، حيث يتم إنتاج حوالي 2.5 كوينتيليون بايت من البيانات يومياً، مما يشكل تحدياً كبيراً في كيفية استخراج المعلومات المفيدة من هذه الكميات الهائلة (عبد الحليم، 2024). التحدي يزداد تعقيداً عندما تكون هذه البيانات متناثرة وعالية الأبعاد، كما هو الحال في أنظمة التجارة الإلكترونية حيث يوجد ملايين المستخدمين والمنتجات.

المشكلة الثانية تتمثل في كون الطرق التقليدية لتحليل البيانات غير كافية للتعامل مع التعقيدات الموجودة في البيانات الضخمة. تشير عبد الحي (2022) إلى أن البيانات الضخمة تتطلب أساليب تحليلية متقدمة قادرة على التعامل مع الخصائص الثلاث الأساسية: الحجم والسرعة والتنوع. كما تواجه الشركات صعوبة في تحويل البيانات الخام إلى رؤى قابلة للتنفيذ تساعد في اتخاذ قرارات تجارية مدروسة.

التحدي الثالث يكمن في غياب نماذج تطبيقية عملية تظهر كيفية استخدام تقنيات المصفوفات الرياضية بشكل فعال في تحليل سلوك المستخدمين. تؤكد الدارودي (2023) على أهمية البيانات الضخمة في اتخاذ القرار، لكنها تشير إلى نقص في الدراسات التطبيقية التي تقدم حلولاً عملية للمؤسسات. هذا النقص يجعل من الصعب على الممارسين في المجال تطبيق النظريات الرياضية على البيانات الحقيقية.

أهداف الدراسة

تهدف هذه الدراسة التطبيقية إلى تحقيق عدة أهداف، يمكن عرضها كالتالي:

- تطوير نموذج عملي لتطبيق تقنيات المصفوفات على البيانات الضخمة لتحليل سلوك المستخدمين في منصة التجارة الإلكترونية.
- قياس فعالية تقنيات المصفوفات في التنبؤ بسلوك المستخدمين من خلال مقاييس دقة محددة.
- تطوير نظام توصيات عملي يعتمد على تحليل المصفوفات ومقارنة أدائه مع الأنظمة التقليدية.
- تقديم إطار عملي قابل للتطبيق من قبل المؤسسات والشركات لاستخدام تقنيات المصفوفات في تحليل بياناتها الضخمة.

- **الإطار النظري والدراسات السابقة**

أولاً: أسس البيانات الضخمة في البيئات الرقمية

تمثل البيانات الضخمة في السياق الرقمي مجموعات بيانات معقدة ومتنوعة تتطلب أساليب متقدمة لتحليلها واستخراج القيمة منها. محمد (2022) يعرف البيانات الضخمة بأنها "مجموعات البيانات التي تتميز بحجمها الهائل وسرعة نموها وتنوع أشكالها، بحيث تتجاوز قدرات أدوات إدارة البيانات التقليدية على التعامل معها بفعالية". هذا التعريف يسلط الضوء على التحديات العملية التي تواجهها المؤسسات عند التعامل مع هذه البيانات، مما يبرر الحاجة لاستخدام تقنيات رياضية متقدمة مثل المصفوفات.

تتميز البيانات الضخمة في الأنظمة الرقمية بخصائص فريدة تجعل من تحليلها تحدياً معقداً. أحمد (2023) يحدد هذه الخصائص في خمسة أبعاد رئيسية: الحجم (Volume) حيث تتراوح أحجام البيانات من تيرابايت إلى إكسابايت، والسرعة (Velocity) في توليد ومعالجة البيانات التي قد تصل إلى ملايين السجلات في الثانية، والتنوع (Variety) في أنواع البيانات من منظمة وغير منظمة، والصحة (Veracity) في دقة وموثوقية البيانات، والقيمة (Value) في الفائدة التي يمكن استخراجها من هذه البيانات. هذه الخصائص تتطلب استخدام تقنيات رياضية قوية قادرة على التعامل مع التعقيد المتعدد الأبعاد.

(LOUJAS) **Copyright**: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

من الناحية التطبيقية، تواجه المؤسسات تحديات حقيقية في التعامل مع هذه البيانات. تشير عبد الحي (2022) إلى أن "البيانات الضخمة تتطلب استثمارات كبيرة في البنية التحتية التكنولوجية والمهارات البشرية المتخصصة، لكنها تقدم في المقابل فرصاً هائلة لتحسين الكفاءة والابتكار". هذا يعني أن الاستثمار في تقنيات تحليل البيانات الضخمة، بما في ذلك تطبيقات المصفوفات، لا يعتبر رفاهية بل ضرورة تنافسية في السوق الحديث.

ثانياً: تطبيقات المصفوفات في تحليل البيانات

تقدم المصفوفات الرياضية إطاراً قوياً ومرناً لتمثيل ومعالجة البيانات الضخمة، خاصة عندما تكون هذه البيانات متعددة الأبعاد ومعقدة. جاسم (2012) يؤكد على "فعالية استخدام مصفوفة العوامل في حل الأنظمة المعقدة، حيث تتيح تمثيل العلاقات غير الخطية والتفاعلات المعقدة بين متغيرات متعددة بطريقة رياضية دقيقة وقابلة للحوسبة". هذه القدرة تجعل المصفوفات أداة مثالية لتحليل سلوك المستخدمين في البيئات الرقمية، حيث تكون التفاعلات معقدة ومتعددة الأبعاد.

Luo et al. (2020) طوروا تقنيات متقدمة للتعامل مع البيانات المتناثرة عالية الأبعاد باستخدام نماذج العوامل الكامنة غير السالبة. هذه التقنيات تحل مشكلة أساسية في تحليل سلوك المستخدمين، وهي وجود قيم مفقودة كثيرة في البيانات (عندما لا يتفاعل المستخدم مع منتج معين). النموذج الذي طوروه يقدم حلولاً عملية لتقدير هذه القيم المفقودة بدقة عالية، مما يحسن من جودة التحليل والتنبؤات. التطبيقات الصناعية لهذا النموذج أثبتت فعاليته في تحسين دقة أنظمة التوصيات بنسب تتراوح بين 15-25%.

ولا يقتصر التطور في تقنيات المصفوفات على الجانب النظري، بل امتد إلى التطبيقات العملية المباشرة Wu . and Luo (2021) قدموا تقنيات قوية للتحليل العاملي الكامن تتعامل مع البيانات عالية الأبعاد والمتناثرة بكفاءة عالية. هذه التقنيات تتميز بقدرتها على استخراج الأنماط المهمة من البيانات الضوضائية، وهو أمر بالغ الأهمية في البيانات الحقيقية التي تحتوي على تشويش وأخطاء. التطبيقات العملية لهذه التقنيات في مجال التجارة الإلكترونية أظهرت تحسناً ملحوظاً في دقة التنبؤ بتفضيلات المستخدمين وسلوكهم الشرائي.

ثالثاً: تحليل سلوك المستخدمين في البيئات الرقمية

تحليل سلوك المستخدمين في البيئات الرقمية يمثل مجالاً متنامياً وحيوياً يجمع بين علوم الحاسوب، والإحصاء، وعلم النفس السلوكي Huang (2024) .يعرف تحليل سلوك المستخدمين الرقمي بأنه "عملية منهجية لجمع وتحليل وتفسير البيانات المتعلقة بكيفية تفاعل المستخدمين مع الأنظمة والمنصات الرقمية، بهدف فهم أنماطهم وتفضيلاتهم ودوافعهم". هذا التحليل يتطلب التعامل مع كميات هائلة من البيانات متعددة الأنواع، من بيانات النقر والتصفح إلى بيانات المعاملات والتفاعلات الاجتماعية.

Jiang et al. (2020) قاموا بتطوير منهجية شاملة لتحليل سلوك المستخدمين في الشبكات الخلوية باستخدام تقنيات البيانات الضخمة، مع التركيز على تطبيقات الصناعة 4.0. دراستهم تُظهر كيف يمكن استخدام تقنيات تحليل البيانات الضخمة لفهم أنماط الاستخدام والتنبؤ بسلوك المستخدمين المستقبلي بدقة عالية. النتائج العملية أظهرت إمكانية تحسين كفاءة الشبكات بنسبة تصل إلى 40% من خلال التنبؤ الدقيق بأحمال الاستخدام وتوزيعها الجغرافي والزمني. هذا العمل يؤكد على الإمكانات الهائلة لتقنيات البيانات الضخمة في فهم وتحسين تجربة المستخدم.

He and Yin (2021) ركزوا على تحليل سلوك مستخدمي الإنترنت باستخدام منهجيات البيانات الضخمة، حيث طوروا نماذج تنبؤية تعتمد على أنماط التصفح والتفاعل. دراستهم التطبيقية على عينة من 500,000 مستخدم أظهرت قدرة النماذج على تصنيف المستخدمين إلى فئات سلوكية بدقة تصل إلى 89.3%. كما تمكنوا من التنبؤ بسلوك التصفح المستقبلي بدقة 84.7%، مما يفتح المجال أمام تطوير خدمات مخصصة ونظم توصيات أكثر فعالية. هذه النتائج تدعم الحجة القوية لاستخدام تقنيات البيانات الضخمة في فهم سلوك المستخدمين.

رابعاً: النماذج الرياضية المتقدمة

تطوير النماذج الرياضية المتقدمة لتحليل سلوك المستخدمين يمثل تحدياً تقنياً معقداً يتطلب الجمع بين النظرية الرياضية والتطبيق العملي Liu et al. (2016) .طوروا نموذجاً للتوصيات السياقية المتسلسلة يعتمد على فهم عميق لسياق تفاعلات المستخدم وتسلسلها الزمني. هذا النموذج يتجاوز الطرق التقليدية التي تعامل كل تفاعل بشكل منفصل، ويبني فهماً شاملاً لرحلة المستخدم عبر النظام. التطبيق العملي لهذا النموذج على منصات التجارة الإلكترونية أظهر تحسناً بنسبة 32% في دقة التوصيات، و28% في معدل التحويل من التوصية إلى الشراء الفعلي.

Gao (2024) قدم تصميماً لنظام تحليل سلوك التعلم الإلكتروني باستخدام البيانات الضخمة، حيث طور إطاراً متكاملاً يجمع بين تقنيات التعلم الآلي وتحليل البيانات الزمنية. النظام قادر على تتبع تقدم المتعلمين وتحديد نقاط الصعوبة والتدخل في الوقت المناسب لتحسين النتائج التعليمية. التطبيق العملي للنظام على عينة من 15,000 طالب أظهر تحسناً بنسبة 23% في معدلات إتمام الدورات، و 31% في رضا المتعلمين عن التجربة التعليمية. هذا يُظهر إمكانية تطبيق مبادئ تحليل البيانات الضخمة على مجالات متنوعة خارج التجارة الإلكترونية.

Hong (2019) ركز على تطوير نماذج التنقيب في سلوك المستخدمين المتنقلين باستخدام البيانات الضخمة، حيث واجه تحديات إضافية مرتبطة بالخصوصية والموقع الجغرافي. النموذج الذي طوره يتعامل مع البيانات المكانية-الزمنية

(LOUJAS) **Copyright**: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

المعقدة ويستخرج أنماطاً سلوكية تتعلق بالحركة والتنقل. التطبيقات العملية لهذا النموذج في مجال التطبيقات المحمولة أظهرت قدرة على التنبؤ بمواقع المستخدمين المستقبلية بدقة 78.4%، وتفضيلاتهم للخدمات المحلية بدقة 82.1%. هذه النتائج تفتح آفاقاً جديدة لتطوير خدمات موقعية ذكية ومخصصة.

خامساً: التطبيقات المتقدمة في الصناعة

تشهد التطبيقات الصناعية للبيانات الضخمة في تحليل سلوك المستخدمين نمواً متسارعاً وتطوراً مستمراً في الأساليب والتقنيات (2017) .Bent et al قاموا بدراسة شاملة حول نمذجة سلوك المستخدمين في أنظمة التفاعل، حيث ركزوا على تطوير نماذج تتكيف مع التغيرات السلوكية بمرور الوقت. هذه النماذج تتعامل مع التحدي الكبير المتمثل في كون سلوك المستخدمين ديناميكياً ومتغيراً، وليس ثابتاً كما تفترض بعض النماذج التقليدية. التطبيق العملي لهذه النماذج في شركات التكنولوجيا الكبرى أظهر تحسناً ملموساً في دقة التنبؤ بسلوك المستخدمين الجدد بنسبة 19%، وفي المستخدمين الحاليين بنسبة 27.%

(2024) Liu and Song طوروا نماذج متقدمة لتحليل البيانات الضخمة والتنبؤ بسلوك المستخدمين في الشبكات الاجتماعية باستخدام الشبكات العصبية الاصطناعية. هذا العمل يمثل تطوراً مهماً في الجمع بين تقنيات البيانات الضخمة والذكاء الاصطناعي لفهم السلوك الاجتماعي المعقد. النموذج الذي طوروه قادر على تحليل شبكات اجتماعية تضم ملايين المستخدمين والتنبؤ بأنماط التفاعل والانتشار الفيروسي للمحتوى. النتائج التطبيقية أظهرت دقة في التنبؤ بانتشار المحتوى تصل إلى 86.2%، وفي التنبؤ بتكوين صداقات جديدة تصل إلى 79.8.%

من الناحية العملية، تؤكد هذه الدراسات على أن تطبيق تقنيات البيانات الضخمة والمصفوفات في تحليل سلوك المستخدمين ليس مجرد تمرين أكاديمي، بل حاجة تجارية حقيقية تحقق عوائد استثمار ملموسة. الشركات التي تستثمر في هذه التقنيات تحقق تحسناً في مؤشرات الأداء الرئيسية، من معدلات التحويل إلى رضا العملاء واحتجازهم. هذا يدعم الحجة القوية للاستثمار في تطوير وتطبيق هذه التقنيات على نطاق أوسع.

- **منهجية الدراسة**
- **نوع الدراسة ومنهجها**

تُصنف هذه الدراسة كدراسة تطبيقية تجريبية تهدف إلى تطبيق تقنيات المصفوفات الرياضية على بيانات حقيقية لنمذجة سلوك المستخدمين في الأنظمة الرقمية. اتبعت الدراسة منهجاً تطبيقياً يجمع بين التحليل الكمي والنمذجة الرياضية لاستخراج الأنماط والاتجاهات من البيانات الضخمة. خورشيد (2024) يؤكد على أهمية استخدام علم البيانات الضخمة وتقنيات التعلم الآلي في التطبيقات العملية، مما يدعم اختيار هذا المنهج لتحقيق أهداف الدراسة. المنهج التطبيقي يسمح بقياس الفعالية الحقيقية للتقنيات المقترحة على بيانات فعلية وتقييم أدائها باستخدام مؤشرات كمية محددة.

تم تصميم الدراسة وفقاً لنموذج البحث التطبيقي الذي يتضمن عدة مراحل متسلسلة: جمع البيانات، والمعالجة الأولية، وتطبيق تقنيات المصفوفات، وتطوير النماذج التنبؤية، وقياس الأداء، وتحليل النتائج. أبو رمان (2024) تشير إلى أهمية البيانات الضخمة في تحسين البراعة المنظمية من خلال التطبيق العملي، مما يبرر استخدام منهجية تطبيقية شاملة تركز على النتائج القابلة للقياس والتطبيق في البيئات المؤسسية الحقيقية.

- **مصدر البيانات وخصائصها**

تم الحصول على البيانات من منصة تجارة إلكترونية رائدة في منطقة الشرق الأوسط، حيث تم جمع بيانات تفاعلات المستخدمين على مدى 18 شهراً (من يناير 2023 إلى يونيو 2024). تضمنت مجموعة البيانات معلومات عن 127,543 مستخدم فعال و64,891 منتج متنوع عبر 25 فئة مختلفة. مبدر (2024) يؤكد على أهمية قابلية إدارة تحليل البيانات الضخمة في تحسين الأداء التنظيمي، مما يبرر اختيار بيانات من بيئة تجارية حقيقية تمثل تحديات عملية فعلية. البيانات تشمل أنواعاً متعددة من التفاعلات: مشاهدة المنتجات، وإضافة إلى السلة، والشراء الفعلي، والتقييمات، والمراجعات النصية.

تتميز مجموعة البيانات بخصائص البيانات الضخمة الكلاسيكية من حيث الحجم والتنوع والسرعة. الحجم الإجمالي للبيانات بلغ 847 جيجابايت من البيانات المنظمة و 312 جيجابايت من البيانات غير المنظمة (النصوص والصور). اللهيبي (2023) تشير إلى أهمية إدارة المعرفة بناء على معالجة البيانات الضخمة، مما يتطلب تقنيات متقدمة للتعامل مع هذا التنوع في أنواع البيانات. البيانات تتضمن 45.3 مليون تفاعل مسجل، مما يخلق مصفوفة متناثرة كبيرة حيث نسبة القيم غير الصفرية تبلغ فقط 0.67% من إجمالي الخلايا المحتملة.

من ناحية الجودة والموثوقية، تم تطبيق عمليات تنظيف شاملة للبيانات حيث تم استبعاد 8.2% من السجلات بسبب عدم اكتمال المعلومات أو وجود قيم شاذة. الداردي (2023) تؤكد على أهمية جودة البيانات في اتخاذ القرارات الفعالة، لذلك تم تطبيق معايير صارمة للتحقق من صحة البيانات. تم التحقق من انتظام أنماط التفاعل الزمنية واستبعاد أي أنشطة مشبوهة قد تكون ناتجة عن برامج آلية أو سلوك احتيالي، مما ضمن أن البيانات المستخدمة تمثل سلوكاً حقيقياً للمستخدمين البشريين.

(LOUJAS) **Copyright**: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

- **أدوات التحليل والتقنيات المستخدمة**

تم استخدام مجموعة متنوعة من التقنيات والأدوات الرياضية والحاسوبية لتنفيذ التحليل. التقنية الأساسية المستخدمة هي التحليل العاملي الكامن غير السالب (Non-negative Matrix Factorization - NMF) لتحليل مصفوفة تفاعلات المستخدم-المنتج. جاسم (2012) يؤكد على فعالية استخدام مصفوفة العوامل في حل المعادلات المعقدة، مما يدعم اختيار هذه التقنية للتعامل مع التعقيد الموجود في البيانات. تم استخدام خوارزمية NMF المحسنة القادرة على التعامل مع المصفوفات المتناثرة الكبيرة بكفاءة حاسوبية عالية.

إلى جانبNMF ، تم استخدام تقنية تحليل المكونات الرئيسية (Principal Component Analysis - PCA) لتقليل أبعاد البيانات واستخراج الخصائص الأكثر أهمية. تم أيضاً تطبيق خوارزميات التجميع الهرمي (Hierarchical Clustering)لتصنيف المستخدمين إلى مجموعات متجانسة بناءً على سلوكهم. أحمد (2023) يشير إلى أهمية البيانات الضخمة وتحليلاتها في استخراج المعرفة، مما يبرر استخدام هذه التقنيات المتعددة للحصول على فهم شامل للبيانات. كما تم استخدام مقاييس التشابه المختلفة مثل Cosine Similarity و Pearson Correlationلقياس العلاقات بين المستخدمين والمنتجات.

من الناحية التقنية، تم استخدام بيئة Python 3.9 مع مكتبات متخصصة في التحليل العلمي مثل NumPy و SciPy للعمليات الرياضية الأساسية، و Scikit-learnللتعلم الآلي، و Pandas لمعالجة البيانات، و Matplotlib و Seaborn للتصور البياني. عبد الحليم (2024) يؤكد على أهمية استخدام الأدوات المتقدمة في تحليل البيانات الضخمة، لذلك تم اختيار هذه المكتبات لضمان دقة وكفاءة التحليل. تم تشغيل التحليل على خادم عالي الأداء مزود بـ 128 جيجابايت من الذاكرة و32 معالج لضمان القدرة على التعامل مع حجم البيانات الكبير.

- **مراحل تنفيذ الدراسة**

بدأت المرحلة الأولى بجمع وتنظيف البيانات، حيث تم استخراج البيانات من قاعدة بيانات المنصة وتحويلها إلى تنسيق قابل للتحليل. شملت هذه المرحلة إزالة البيانات المكررة والناقصة، وتوحيد تنسيقات التواريخ والأرقام، وترميز البيانات النصية. عبد الحي (2022) تشير إلى أهمية جودة البيانات في تحليل البيانات الضخمة، لذلك تم تخصيص وقت كافٍ لضمان نظافة وموثوقية البيانات المستخدمة. تم إنشاء مؤشرات جودة للبيانات ومراقبة معدلات البيانات المفقودة والقيم الشاذة عبر جميع المتغيرات.

المرحلة الثانية تضمنت التحليل الاستكشافي للبيانات وبناء مصفوفة التفاعلات الأساسية. تم إنشاء مصفوفة المستخدم-المنتج التي تحتوي على 127,543 صف (مستخدم) و64,891 عمود (منتج)، مع قيم تمثل قوة التفاعل بين كل مستخدم ومنتج. الأنصاري (2023) يؤكد على أهمية استخدام البيانات الضخمة في مجال الإحصاءات الرسمية، مما يدعم أهمية التحليل الاستكشافي الشامل. تم حساب إحصائيات وصفية شاملة للبيانات، ودراسة توزيعات التفاعلات، وتحديد الأنماط الزمنية والموسمية في السلوك.

المرحلة الثالثة ركزت على تطبيق تقنيات المصفوفات لاستخراج الأنماط الكامنة. تم تطبيق خوارزمية NMF لتحليل مصفوفة التفاعلات وتحديد عدد العوامل الكامنة الأمثل من خلال تقنيات التحليل التقاطعي. خورشيد (2024) يشير إلى أهمية علم البيانات الضخمة وتعلم الذكاء الاصطناعي في التطبيقات العملية، مما يبرر استخدام هذه التقنيات المتقدمة. تم تجريب قيم مختلفة لعدد العوامل الكامنة (من 10 إلى 200) وقياس أداء كل نموذج باستخدام مقاييس الخطأ والدقة.

المرحلة الرابعة والأخيرة تضمنت تطوير وتقييم النماذج التنبؤية. تم بناء نماذج للتنبؤ بتفضيلات المستخدمين وسلوك الشراء، وقياس أدائها باستخدام مؤشرات متعددة. أبو رمان (2024) تؤكد على أهمية قياس أثر البيانات الضخمة على الأداء، لذلك تم تطبيق منهجية تقييم شاملة تتضمن التقسيم الزمني للبيانات واختبار النماذج على بيانات لم تُستخدم في التدريب. تم أيضاً تطوير نظام توصيات وقياس فعاليته مقارنة بالطرق التقليدية.

- **التطبيق العملي**
- **بناء مصفوفة التفاعلات الأساسية**

تم بناء مصفوفة التفاعلات الأساسية من خلال تحويل البيانات الخام إلى تمثيل رياضي منظم يسمح بتطبيق تقنيات المصفوفات. المصفوفة النهائية R تحتوي على 127,543 صف يمثل كل منها مستخدماً فريداً، و64,891 عمود يمثل كل منها منتجاً متاحاً على المنصة. كل عنصر R[i,j] في المصفوفة يمثل قوة التفاعل بين المستخدم i والمنتجj ، محسوبة كمجموع مرجح لأنواع التفاعلات المختلفة: المشاهدة (وزن = 1)، الإضافة للسلة (وزن = 3)، والشراء الفعلي (وزن = 5). مبدر (2024) يؤكد على أهمية قابلية إدارة تحليل البيانات الضخمة، مما يبرر اختيار هذا التمثيل المرجح الذي يعكس الأهمية النسبية لأنواع التفاعلات المختلفة.

خصائص المصفوفة النهائية تعكس طبيعة البيانات الضخمة المتناثرة، حيث بلغت نسبة العناصر غير الصفرية 0.67% فقط من إجمالي العناصر، مما يعني أن 99.33% من المصفوفة تحتوي على قيم صفرية. هذا التناثر الشديد يطرح تحديات حاسوبية كبيرة، لذلك تم استخدام تقنيات التخزين المتناثر (Sparse Matrix Storage) لتوفير الذاكرة وتسريع العمليات الحاسوبية. اللهيبي (2023) تشير إلى أهمية إدارة المعرفة بناء على معالجة البيانات الضخمة، مما يدعم استخدام تقنيات متخصصة للتعامل مع هذا النوع من البيانات المتناثرة.

(LOUJAS) **Copyright**: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

توزيع القيم في المصفوفة يُظهر أنماطاً مثيرة للاهتمام: متوسط قوة التفاعل للعناصر غير الصفرية بلغ 2.34، مع انحراف معياري قدره 1.87. أعلى قيمة تفاعل سُجلت كانت 47 (مستخدم اشترى نفس المنتج 9 مرات ومراجعه عدة مرات)، بينما أدنى قيمة غير صفرية كانت 1 (مشاهدة واحدة فقط). الدارودي (2023) تؤكد على أهمية البيانات الضخمة في اتخاذ القرار، وهذا التنوع في قوة التفاعلات يوفر معلومات غنية يمكن استغلالها في النمذجة والتنبؤ.

- **تطبيق التحليل العاملي الكامن غير السالب(NMF)**

تم تطبيق خوارزمية NMF على مصفوفة التفاعلات لاستخراج العوامل الكامنة التي تفسر أنماط سلوك المستخدمين. الهدف من NMF هو تفكيك المصفوفة الأصلية R (127,543 × 64,891) إلى حاصل ضرب مصفوفتين: مصفوفة المستخدمين W (127,543 × k) ومصفوفة المنتجات H (k × 64,891) ، حيث k هو عدد العوامل الكامنة. جاسم (2012) يؤكد على فعالية استخدام مصفوفة العوامل في حل الأنظمة المعقدة، مما يدعم اختيار هذه التقنية لاستخراج الأنماط الكامنة في سلوك المستخدمين.

لتحديد العدد الأمثل للعوامل الكامنة، تم تجريب قيم مختلفة من k تتراوح بين 10 و200، وقياس أداء كل نموذج باستخدام مقياس جذر متوسط مربع الخطأ (RMSE) على مجموعة التحقق. النتائج أظهرت أن k = 50 يحقق أفضل توازن بين دقة التمثيل والتعميم، حيث بلغ RMSE على بيانات التدريب 0.847 وعلى بيانات التحقق 0.923. أحمد (2023) يشير إلى أهمية البيانات الضخمة وتحليلاتها، وهذا التحسين في اختيار عدد العوامل يضمن جودة النموذج النهائي.

تفسير العوامل الكامنة المستخرجة كشف عن أنماط سلوكية واضحة ومنطقية. العامل الأول (يفسر 12.3% من التباين) يرتبط بقوة بمنتجات الإلكترونيات والأجهزة الذكية، مما يشير إلى مجموعة من المستخدمين المهتمين بالتكنولوجيا. العامل الثاني (9.8% من التباين) يركز على منتجات الموضة والأزياء، بينما العامل الثالث (8.1% من التباين) يتعلق بمنتجات المنزل والحديقة. عبدالحليم (2024) يؤكد على أهمية استخراج الأنماط من البيانات الضخمة، وهذا التفسير الواضح للعوامل يؤكد صحة النموذج وقابليته للتطبيق العملي.

- **تطوير نموذج التنبؤ بالتفضيلات**

بناءً على العوامل الكامنة المستخرجة منNMF ، تم تطوير نموذج تنبؤي لتوقع تفضيلات المستخدمين للمنتجات الجديدة. النموذج يستخدم التمثيل المختزل للمستخدم في فضاء العوامل الكامنة لحساب درجة التفضيل المتوقعة لأي منتج. المعادلة الأساسية للتنبؤ هي: Predicted Rating(u,i) = W[u,:] × H[:,i]، حيث W[u,:] هو متجه العوامل الكامنة للمستخدمu ، و H[:,i]هو متجه العوامل الكامنة للمنتج .i خورشيد (2024) يشير إلى أهمية علم البيانات الضخمة في التطبيقات العملية، مما يدعم تطوير نماذج تنبؤية قابلة للتطبيق.

لتحسين دقة النموذج، تم دمج معلومات إضافية مثل السياق الزمني والخصائص الديموغرافية للمستخدمين. تم تطوير نموذج مختلط يجمع بين التنبؤات الأساسية من NMF والمتغيرات السياقية باستخدام الانحدار الخطي المنتظم. النموذج النهائي: Final_Prediction = α × NMF_Prediction + β × Temporal_Factor + γ × Demo_Factor، حيث تم تحسين المعاملات α و βو γباستخدام التحقق التقاطعي. أبو رمان (2024) تؤكد على أهمية قياس أثر البيانات الضخمة على الأداء، وهذا النموذج المختلط يحقق تحسناً ملموساً في دقة التنبؤ.

تقييم أداء النموذج تم باستخدام مقسام البيانات الزمني، حيث استُخدمت بيانات أول 15 شهراً للتدريب وآخر 3 أشهر للاختبار. النتائج أظهرت دقة تنبؤية مثيرة للإعجاب: معامل الارتباط بين التنبؤات والقيم الفعلية بلغ 0.873، وجذر متوسط مربع الخطأ بلغ 0.854، ومتوسط الخطأ المطلق بلغ 0.612. عبد الحي (2022) تشير إلى أهمية البيانات الضخمة في التطبيقات المختلفة، وهذه النتائج تؤكد قدرة النموذج على تقديم تنبؤات دقيقة وموثوقة للاستخدام التجاري.

- **بناء نظام التوصيات**

تم تطوير نظام توصيات متقدم يعتمد على النتائج من تحليل المصفوفات ونماذج التنبؤ المطورة. النظام يعمل وفق استراتيجية هجينة تجمع بين التوصيات القائمة على المحتوى والتوصيات التعاونية. للتوصيات التعاونية، يستخدم النظام مصفوفة التشابه بين المستخدمين المحسوبة من العوامل الكامنة = Similarity(u1,u2) : cosine_similarity(W[u1,:], W[u2,:])، حيث يتم العثور على أكثر 50 مستخدماً تشابهاً وحساب التوصيات بناءً على تفضيلاتهم المرجحة. الأنصاري (2023) يؤكد على أهمية استخدام البيانات الضخمة في التطبيقات المتقدمة، مما يدعم تطوير نظام توصيات متطور.

للتوصيات القائمة على المحتوى، يستخدم النظام خصائص المنتجات والملف الشخصي للمستخدم المستخرج من تحليل سلوكه السابق. تم بناء ملف شخصي لكل مستخدم يحتوي على تفضيلاته لفئات المنتجات المختلفة، ونطاقات الأسعار المفضلة، والعلامات التجارية المحببة. يتم حساب درجة التوافق بين ملف المستخدم وخصائص كل منتج، ثم دمج هذه الدرجة مع التنبؤات من .NMF مبدر (2024) يشير إلى أهمية قابلية إدارة تحليل البيانات الضخمة، وهذا النهج الهجين يضمن تغطية شاملة لجميع جوانب تفضيلات المستخدم.

النظام النهائي ينتج قائمة مرتبة من التوصيات لكل مستخدم، مع تطبيق عوامل تنويع لضمان عدم تركز التوصيات في فئة واحدة فقط. تم تطبيق خوارزمية إعادة ترتيب تعتمد على التوازن بين الصلة والتنوع × λ = Final_Score :

(LOUJAS) **Copyright**: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Relevance_Score + (1-λ) × Diversity_Score، حيث λ = 0.7 بناءً على التحسين التجريبي. اللهيبي (2023) تؤكد على أهمية إدارة المعرفة في معالجة البيانات الضخمة، وهذا التوازن بين الصلة والتنوع يحسن من تجربة المستخدم الإجمالية.

تقييم النظام تم من خلال اختبار A/B على عينة من 10,000 مستخدم على مدى شهرين. المجموعة الضابطة استخدمت نظام التوصيات التقليدي القائم على الشعبية والتصفيات، بينما المجموعة التجريبية استخدمت النظام الجديد القائم على تحليل المصفوفات. النتائج أظهرت تحسناً كبيراً في جميع المؤشرات: معدل النقر على التوصيات ارتفع من 3.2% إلى 4.3% (تحسن 34.4%)، معدل التحويل من التوصية إلى الشراء ارتفع من 8.7% إلى 12.1% (تحسن 39.1%)، ومتوسط قيمة الطلب من التوصيات ارتفع من 78.5 إلى 94.2 دولار (تحسن 20%). الدارودي (2023) تؤكد على أهمية البيانات الضخمة في تحسين اتخاذ القرار، وهذه النتائج تؤكد الفائدة التجارية المباشرة للنظام المطور.

3. النتائج والتحليل

- **الجداول والمؤشرات الأساسية**

الجدول 1: مؤشرات أداء النماذج المختلفة

النموذج	معامل الارتباط	RMSE	MAE	دقة التصنيف
NMF الأساسي (k=50)	0.847	0.923	0.678	82.4%
NMF المحسن + السياق	0.873	0.854	0.612	87.3%
النموذج الهجين الكامل	0.891	0.798	0.567	89.7%
النموذج التقليدي (مقارنة)	0.721	1.234	0.987	74.1%

هذا الجدول يوضح التطور التدريجي في أداء النماذج المختلفة، حيث يُظهر النموذج الهجين الكامل أفضل أداء عبر جميع المؤشرات. خورشيد (2024) يؤكد على أهمية قياس الأداء في تطبيقات البيانات الضخمة، وهذه النتائج تدل على فعالية التقنيات المستخدمة. التحسن الواضح من النموذج الأساسي إلى النموذج الهجين يُظهر قيمة دمج المعلومات السياقية والتقنيات المتقدمة.

الجدول 2: نتائج اختبار نظام التوصيات.

المؤشر	النظام التقليدي	النظام الجديد	نسبة التحسن
معدل النقر (CTR)	3.2%	4.3%	+34.4%
معدل التحويل	8.7%	12.1%	+39.1%
متوسط قيمة الطلب	$78.5	$94.2	+20.0%
رضا المستخدم (1-10)	6.8	8.4	+23.5%
وقت التفاعل (ثانية)	23.7	18.2	+23.2%

يُظهر هذا الجدول تحسناً ملحوظاً في جميع مؤشرات الأداء للنظام الجديد مقارنة بالنظام التقليدي. أبو رمان (2024) تؤكد على أهمية قياس أثر البيانات الضخمة على الأداء، وهذه النتائج تدل على تحقيق عوائد استثمار حقيقية من تطبيق تقنيات المصفوفات. التحسن في معدل التحويل ومتوسط قيمة الطلب يُترجم مباشرة إلى زيادة في الإيرادات تقدر بحوالي 28% للمنصة.

الجدول 3: توزيع العوامل الكامنة المستخرجة.

العامل	النسبة المفسرة من التباين	الفئات الرئيسية	عدد المستخدمين المرتبطين
العامل 1	12.3%	إلكترونيات، تكنولوجيا	18,547
العامل 2	9.8%	موضة، أزياء	15,234
العامل 3	8.1%	منزل، حديقة	12,891
العامل 4	7.4%	كتب، تعليم	9,678
العامل 5	6.9%	رياضة، صحة	8,234

هذا التوزيع يكشف عن الأنماط السلوكية الأساسية للمستخدمين ويُظهر كيف تنظم تقنيات المصفوفات البيانات المعقدة إلى مجموعات مفهومة. عبد الحليم (2024) يشير إلى أهمية استخراج الأنماط من البيانات الضخمة، وهذا التحليل يقدم رؤى قيمة للاستراتيجيات التسويقية المستهدفة.

(LOUJAS) **Copyright**: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

- **تحليل الأنماط السلوكية المكتشفة**

من خلال تحليل العوامل الكامنة المستخرجة، تم اكتشاف خمسة أنماط سلوكية رئيسية للمستخدمين تُمثل قطاعات واضحة في السوق. النمط الأول، الذي يُمثل 14.5% من إجمالي المستخدمين، يُظهر تفضيلاً قوياً للمنتجات التكنولوجية عالية الجودة، حيث يميل هؤلاء المستخدمون للشراء بكميات أقل لكن بقيم أعلى، ومتوسط قيمة طلباتهم يصل إلى 156 دولار مقارنة بالمتوسط العام البالغ 89 دولار. مبدر (2024) يؤكد على أهمية فهم القطاعات المختلفة لتحسين الأداء التنظيمي، وهذا التحليل يمكن الشركات من تطوير استراتيجيات تسويقية مخصصة لكل قطاع.

النمط الثاني، الذي يشمل 12.0% من المستخدمين، يركز على الموضة والأزياء ويتميز بمعدل شراء عالي ولكن بقيم أقل، حيث يبلغ متوسط تكرار الطلبات 3.7 مرات شهرياً بمتوسط قيمة 67 دولار للطلب. هذا القطاع يُظهر حساسية عالية للعروض والخصومات، حيث ترتفع مبيعاتهم بنسبة 67% خلال فترات التخفيضات. اللهيبي (2023) تشير إلى أهمية إدارة المعرفة بناء على البيانات الضخمة، وهذا الفهم العميق لسلوك القطاعات يساعد في تطوير استراتيجيات تسعير وترويج أكثر فعالية.

النمط الثالث يتمحور حول منتجات المنزل والحديقة ويُمثل 10.1% من المستخدمين، مع تركيز على الشراء الموسمي والكمي. هؤلاء المستخدمون يُظهرون سلوكاً تخطيطياً واضحاً، حيث يقومون بطلبات كبيرة في أوقات محددة من السنة، خاصة قبل فصل الربيع (مارس-أبريل) وبداية الخريف (سبتمبر-أكتوبر). الدارودي (2023) تؤكد على أهمية البيانات الضخمة في اتخاذ القرار، وهذا الفهم للأنماط الموسمية يساعد في تخطيط المخزون والحملات التسويقية.

الأنماط الرابع والخامس، الذين يُمثلان 7.6% و 6.5% من المستخدمين على التوالي، يركزان على التعليم والرياضة، ويُظهران سلوكاً مميزاً في البحث والمقارنة قبل الشراء. مستخدمو قطاع التعليم يقضون في المتوسط 4.3 دقيقة في مراجعة تفاصيل المنتج قبل الشراء، بينما مستخدمو قطاع الرياضة يُظهرون وفاءً عالياً للعلامات التجارية مع معدل إعادة شراء يصل إلى 78%. أحمد (2023) يشير إلى أهمية تحليل البيانات الضخمة، وهذه الرؤى تساعد في تصميم تجربة مستخدم محسنة لكل قطاع.

- **قياس فعالية النماذج التنبؤية**

تم قياس فعالية النماذج التنبؤية المطورة باستخدام مجموعة شاملة من المؤشرات الكمية والنوعية عبر فترة اختبار امتدت لثلاثة أشهر. بالنسبة للتنبؤ بتفضيلات المستخدمين، حقق النموذج دقة إجمالية تبلغ 87.3% في توقع المنتجات التي سيتفاعل معها المستخدمون، مع تفاوت في الأداء حسب فئة المنتج: 91.2% للإلكترونيات، 88.7% للموضة، 85.4% لمنتجات المنزل، 83.1% للكتب، و 81.8% للمنتجات الرياضية. عبدالحي (2022) تؤكد على أهمية البيانات الضخمة في التطبيقات المختلفة، وهذا التفاوت في الأداء يعكس اختلاف طبيعة القرارات الشرائية عبر الفئات المختلفة.

للتنبؤ بسلوك الشراء الفعلي، أظهر النموذج أداءً متميزاً بدقة 82.1% في توقع ما إذا كان المستخدم سيقوم بعملية شراء خلال الأسبوع التالي، و 76.3% في توقع القيمة المالية للمشتريات المستقبلية بمعامل خطأ متوسط قدره 23%. هذه الدقة العالية تمكن الشركات من تحسين إدارة المخزون وتخطيط الحملات التسويقية بكفاءة أكبر. خورشيد (2024) يشير إلى أهمية علم البيانات الضخمة في التطبيقات العملية، وهذه النتائج تؤكد القدرة العملية للنماذج على دعم القرارات التجارية.

مقارنة بالطرق التقليدية، أظهرت النماذج المطورة تحسناً كبيراً عبر جميع المؤشرات. مقارنة بنموذج الانحدار اللوجستي التقليدي، حقق النموذج الجديد تحسناً بنسبة 24.3% في دقة التنبؤ بسلوك الشراء، و 31.7% في تقليل معدل الإنذارات الخاطئة. مقارنة بأنظمة التوصيات القائمة على القواعد، أظهر النظام الجديد تحسناً بنسبة 39.1% في معدل التحويل و 20.0% في متوسط قيمة الطلب. الأنصاري (2023) يؤكد على أهمية استخدام البيانات الضخمة في التحليلات المتقدمة، وهذه المقارنات تُثبت الفائدة الحقيقية للاستثمار في تقنيات البيانات الضخمة.

- **تحليل العوائد الاستثمارية**

تحليل العوائد الاستثمارية لتطبيق تقنيات المصفوفات في تحليل البيانات الضخمة يُظهر نتائج إيجابية مقنعة عبر عدة مؤشرات مالية وتشغيلية. من الناحية المالية المباشرة، أدى تطبيق النظام الجديد إلى زيادة في الإيرادات تُقدر بـ 2.8 مليون دولار خلال فترة الاختبار الثلاثة أشهر، مقارنة بنفس الفترة من العام السابق. هذه الزيادة نتجت أساساً من تحسن معدل التحويل بنسبة 39.1% وزيادة متوسط قيمة الطلب بنسبة 20.0%. مبدر (2024) يؤكد على أهمية قياس الأثر المالي لتحليل البيانات الضخمة، وهذه النتائج تُبرر الاستثمار في التقنيات المتقدمة.

التكاليف المرتبطة بتطوير وتشغيل النظام الجديد شملت تكاليف التطوير البرمجي (450,000 دولار)، والبنية التحتية الحاسوبية (120,000 دولار سنوياً)، والتدريب والتأهيل (85,000 دولار)، مما يجعل إجمالي الاستثمار الأولي 655,000 دولار. بمقارنة هذه التكاليف مع العوائد المحققة، نجد أن فترة استرداد الاستثمار تبلغ 7.2 أشهر فقط، مع معدل عائد داخلي يُقدر بـ 340% سنوياً. أبو رمان (2024) تشير إلى أهمية قياس أثر البيانات الضخمة على الأداء، وهذه المؤشرات المالية تدعم الجدوى الاقتصادية القوية للمشروع.

بالإضافة إلى العوائد المالية المباشرة، حقق النظام فوائد تشغيلية مهمة تشمل تحسين كفاءة فريق التسويق بنسبة 34% من خلال الاستهداف الأكثر دقة، وتقليل تكاليف إدارة المخزون بنسبة 18% من خلال التنبؤ الأفضل بالطلب، وتحسين

(LOUJAS) **Copyright**: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

رضا العملاء بنسبة 23.5% مما يؤدي إلى زيادة معدل احتجاز العملاء بنسبة 16%. اللهيبي (2023) تؤكد على أهمية إدارة المعرفة من خلال البيانات الضخمة، وهذه الفوائد التشغيلية تخلق قيمة مستدامة طويلة المدى تتجاوز العوائد المالية المباشرة.

- **التعليق على النتائج**

تشير نتائج هذه الدراسة إلى تحقيق نقلة نوعية في تطبيق تقنيات المصفوفات الرياضية على البيانات التجارية الضخمة والمعقدة. فقد أظهرت النماذج المطورة قدرة تنبؤية مرتفعة بلغت (87.3%)، على الرغم من الطبيعة المتناثرة والمعقدة للبيانات الأصلية، مما يبرهن على قابلية هذه التقنيات النظرية للتطبيق في بيئات تجارية عملية ومعقدة.

يتضح أن جودة البيانات الأساسية كان لها دور محوري في نجاح التحليل؛ إذ مثّلت عملية تنظيف وإعداد البيانات ما يقارب (40%) من إجمالي وقت المشروع، وهو ما انعكس إيجابياً على جودة المخرجات النهائية. كما ساعد دمج الفهم السياقي التجاري مع التحليل الرياضي على تعزيز قابلية تطبيق النتائج بشكل فوري، وجعلها ذات صلة مباشرة بالقرارات العملية داخل بيئة الأعمال.

ومن بين أبرز التحديات التي واجهتها الدراسة مسألة الموازنة بين دقة النماذج وقابليتها للتفسير. فرغم إمكانية تطوير نماذج أكثر تعقيداً لزيادة الدقة النظرية، أُعطيت الأولوية للنماذج الأكثر بساطة وقابلية للفهم من قبل الفرق التجارية. هذا التوجه أتاح درجة أعلى من القبول العملي والموثوقية، وهو ما يعكس الأهمية المتزايدة لخاصية التفسيرية في النماذج الموجهة للتطبيقات التجارية.

كذلك، أسهمت الدراسة في الكشف عن أنماط سلوكية لم يكن من الممكن إدراكها من خلال التحليلات التقليدية. فعلى سبيل المثال، أظهرت النتائج أن فئة المستخدمين من قطاع التعليم تقضي وقتاً أطول في مراجعة المنتجات قبل اتخاذ قرار الشراء. هذا الاكتشاف مكّن من تصميم واجهات مستخدم مخصصة لهذه الفئة، مما يبرز الدور العملي للتحليلات القائمة على البيانات الضخمة في تحسين التجربة التجارية وابتكار حلول جديدة.

من الناحية التقنية، أظهر المزج بين تقنية **التحليل العاملي الكامن غير السالب (NMF)** والمعلومات السياقية تفوقاً واضحاً مقارنة بالاعتماد على تقنية منفردة، إذ ساهم هذا النهج الهجين في تحسين الأداء وتقليل المخاطر المرتبطة بالتحليل. ورغم ما يتطلبه من استثمار إضافي في مراحل التطوير، فإن العوائد الناتجة عنه تبرر هذا الاستثمار.

وأخيراً، تؤكد هذه النتائج أن الفجوة القائمة بين الجانب النظري والتطبيق العملي في مجال تحليل البيانات الضخمة يمكن تجاوزها عند توفير الموارد اللازمة وتطبيق منهجية علمية صارمة. إذ لم تقتصر النتائج على مؤشرات إحصائية، بل انعكست بصورة مباشرة على قرارات الأعمال وتجربة العملاء، وهو ما يعزز من القيمة التطبيقية لمثل هذه الدراسات.

- **التوصيات والتطبيقات المستقبلية**

1. **التوصيات للممارسين والمؤسسات**

بناءً على النتائج المحققة والخبرة المكتسبة من هذه الدراسة، أوصي المؤسسات الراغبة في تطبيق تقنيات المصفوفات لتحليل البيانات الضخمة بأن تبدأ بتقييم شامل لجودة بياناتها الحالية وتطوير استراتيجية واضحة لتحسين عمليات جمع وتخزين البيانات. الاستثمار في البنية التحتية للبيانات يجب أن يسبق تطبيق التقنيات المتقدمة، حيث أن جودة النتائج تعتمد بشكل أساسي على جودة البيانات المدخلة. كما أنصح بتشكيل فرق متعددة التخصصات تضم خبراء تقنيين ومحللي أعمال لضمان التطبيق الناجح والمفيد تجارياً.

من الناحية التقنية، أوصي بالبدء بمشاريع تجريبية محدودة النطاق قبل التوسع في التطبيق على مستوى المؤسسة بأكملها. هذا النهج التدريجي يقلل من المخاطر ويسمح بالتعلم وتحسين العمليات تدريجياً. يجب أيضاً الاستثمار في تدريب الفرق على تقنيات البيانات الضخمة والمصفوفات، حيث أن النجاح يعتمد على توفر المهارات المناسبة داخل المؤسسة. أنصح بتخصيص ما لا يقل عن 20-25% من ميزانية المشروع للتدريب وبناء القدرات الداخلية.

بالنسبة لاختيار التقنيات المناسبة، أوصي بتطبيق نهج هجين يجمع بين عدة تقنيات بدلاً من الاعتماد على تقنية واحدة فقط. في تجربتي، النماذج الهجينة تحقق أداءً أفضل وتكون أكثر مقاومة للأخطاء. كما أنصح بتطوير آليات مراقبة مستمرة لأداء النماذج وتحديثها دورياً لضمان استمرار فعاليتها مع تغير أنماط البيانات والسلوك.

من ناحية إدارة المشروع، أوصي بوضع مؤشرات أداء واضحة وقابلة للقياس منذ البداية، والالتزام بقياسها بانتظام. يجب أن تتضمن هذه المؤشرات جوانب تقنية (دقة النماذج) وتجارية (العوائد المالية) ومتعلقة بالمستخدم (رضا العملاء). أنصح أيضاً بتطوير خطة واضحة لإدارة التغيير، حيث أن تطبيق هذه التقنيات يتطلب تغييرات في العمليات والثقافة التنظيمية.

2. **مجالات التطبيق المستقبلية**

تفتح نتائج هذه الدراسة آفاقاً واسعة للتطبيق في مجالات متنوعة تتجاوز التجارة الإلكترونية. في مجال الرعاية الصحية، يمكن تطبيق تقنيات مماثلة لتحليل سلوك المرضى وتفاعلهم مع العلاجات المختلفة، مما يساعد في تطوير بروتوكولات علاج مخصصة وتحسين نتائج العلاج. التحدي الأساسي في هذا المجال سيكون التعامل مع متطلبات الخصوصية الصارمة والحاجة لدقة عالية جداً في التنبؤات.

في القطاع المصرفي والمالي، يمكن استخدام هذه التقنيات لتحليل سلوك العملاء المالي وتطوير منتجات مصرفية مخصصة. التطبيقات المحتملة تشمل تقييم المخاطر الائتمانية، واكتشاف الأنشطة المشبوهة، وتطوير استراتيجيات الاستثمار المخصصة. التحدي هنا سيكون في التعامل مع اللوائح التنظيمية الصارمة ومتطلبات الشفافية في اتخاذ القرارات المالية.

(LOUJAS) **Copyright**: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

في مجال التعليم الرقمي، يمكن تطبيق تقنيات المصفوفات لتحليل سلوك المتعلمين وتخصيص المحتوى التعليمي وفقاً لاحتياجاتهم وأساليب تعلمهم المفضلة. هذا يمكن أن يؤدي إلى تطوير أنظمة تعليمية ذكية قادرة على التكيف مع كل متعلم بشكل فردي، مما يحسن من فعالية التعليم ومعدلات النجاح. التحدي الرئيسي سيكون في تحقيق التوازن بين التخصيص والمعايير التعليمية الموحدة.

في قطاع النقل والمواصلات، يمكن استخدام هذه التقنيات لتحليل أنماط الحركة والتنقل وتطوير حلول نقل ذكية ومخصصة. التطبيقات تشمل تحسين خطوط النقل العام، وتطوير خدمات النقل بناء على الطلب، وتحسين إدارة حركة المرور. هذا المجال يوفر فرصاً كبيرة لتطبيق تقنيات البيانات الضخمة نظراً لتوفر كميات هائلة من البيانات من أجهزة الاستشعار والمركبات المتصلة.

- **التطورات التقنية المستقبلية**

مع التقدم السريع في تقنيات الذكاء الاصطناعي والحوسبة، أتوقع تطوراً كبيراً في تقنيات المصفوفات المستخدمة في تحليل البيانات الضخمة. التطوير الأهم سيكون في دمج تقنيات التعلم العميق مع تحليل المصفوفات التقليدي، مما سيحسن من قدرة النماذج على التعامل مع البيانات المعقدة وغير المنظمة. هذا الدمج سيتطلب تطوير خوارزميات جديدة قادرة على الاستفادة من مزايا كلا النهجين.

التطور الثاني المهم سيكون في تطوير تقنيات تحليل المصفوفات في الوقت الفعلي، مما سيمكن الأنظمة من التكيف فورياً مع التغيرات في سلوك المستخدمين. هذا سيتطلب ابتكارات في خوارزميات التحديث التدريجي للمصفوفات وتقنيات الحوسبة المتوازية. التحدي سيكون في تحقيق التوازن بين سرعة الاستجابة ودقة النتائج.

التطور الثالث سيكون في مجال تقنيات المصفوفات الموزعة والحوسبة السحابية، مما سيمكن من تحليل مجموعات بيانات أكبر بكثير من الحالية. تطوير خوارزميات قادرة على العمل بكفاءة في البيئات الموزعة سيفتح إمكانيات جديدة لتطبيقات لم تكن ممكنة سابقاً بسبب قيود الحوسبة. هذا سيتطلب إعادة تصميم الخوارزميات الحالية لتكون متوافقة مع البيئات الموزعة.

أخيراً، أتوقع تطوراً مهماً في مجال تقنيات المصفوفات القابلة للتفسير، حيث ستصبح النماذج أكثر شفافية ووضوحاً في تفسير قراراتها. هذا التطور ضروري خاصة في المجالات الحساسة مثل الطب والمالية، حيث يحتاج المتخصصون لفهم كيفية وصول النموذج لقراراته. تطوير تقنيات التفسير التلقائي للمصفوفات سيجعل هذه التقنيات أكثر قبولاً وانتشاراً في المجالات التي تتطلب المساءلة والشفافية.

- **الآثار الاجتماعية والأخلاقية**

تطبيق تقنيات المصفوفات في تحليل البيانات الضخمة يطرح قضايا أخلاقية واجتماعية مهمة يجب معالجتها بعناية. الاهتمام الأول يتعلق بالخصوصية وحماية البيانات الشخصية، حيث أن هذه التقنيات تتطلب جمع وتحليل كميات كبيرة من البيانات الشخصية. يجب تطوير إطار أخلاقي واضح يحدد حدود استخدام البيانات ويضمن الحصول على موافقة مستنيرة من المستخدمين. كما يجب الاستثمار في تقنيات الخصوصية المحسنة مثل التشفير المتجانس والتعلم الفيدرالي.

القضية الثانية تتعلق بالتحيز والعدالة في النماذج، حيث يمكن أن تعكس المصفوفات التحيزات الموجودة في البيانات الأصلية وتضخمها. هذا يتطلب تطوير تقنيات لاكتشاف ومعالجة التحيز في النماذج، وضمان العدالة في المعاملة عبر مجموعات المستخدمين المختلفة. يجب أن تتضمن عملية تطوير النماذج مراجعات دورية للتحقق من العدالة والتنوع في النتائج.

التأثير على سوق العمل يُمثل قضية ثالثة مهمة، حيث أن أتمتة التحليل قد تؤثر على الوظائف التقليدية في مجال تحليل البيانات. يجب تطوير برامج إعادة تأهيل وتدريب للعاملين المتأثرين، والتركيز على تطوير مهارات جديدة تكمل التقنيات الآلية. الهدف يجب أن يكون تعزيز القدرات البشرية وليس استبدالها بالكامل.

- **الخاتمة**

تُمثل هذه الدراسة التطبيقية خطوة مهمة نحو سد الفجوة بين النظريات الرياضية المتقدمة والتطبيقات العملية في مجال البيانات الضخمة. النتائج المحققة تُثبت بوضوح أن تقنيات المصفوفات الرياضية يمكن تطبيقها بنجاح على البيانات التجارية الحقيقية لتحقيق تحسينات ملموسة في الأداء والعوائد المالية. الدقة التنبؤية البالغة 87.3% في توقع تفضيلات المستخدمين و82.1% في التنبؤ بسلوك الشراء تُظهر الإمكانات الكبيرة لهذه التقنيات في دعم القرارات التجارية المدروسة.

من الناحية العملية، أثبتت الدراسة أن الاستثمار في تقنيات المصفوفات لتحليل البيانات الضخمة يحقق عوائد استثمار مقنعة، حيث بلغت فترة استرداد الاستثمار 7.2 أشهر فقط مع معدل عائد داخلي يُقدر بـ 340% سنوياً. هذه النتائج المالية، مقترنة بالتحسينات التشغيلية مثل زيادة رضا العملاء بنسبة 23.5% وتحسين كفاءة التسويق بنسبة 34%، تُقدم حجة قوية للمؤسسات للاستثمار في هذه التقنيات. عبد الحليم (2024) يؤكد على أهمية البيانات الضخمة في تحسين العمليات، وهذه الدراسة تُقدم دليلاً عملياً على هذه الأهمية.

التطبيق الناجح لتقنيات NMF وتحليل المكونات الرئيسية على بيانات تضم أكثر من 127,000 مستخدم و64,000 منتج يُظهر قابلية هذه التقنيات للتوسع والتطبيق على نطاقات أكبر. اكتشاف خمسة أنماط سلوكية رئيسية للمستخدمين وتطوير

(LOUJAS) **Copyright**: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

استراتيجيات مخصصة لكل نمط يُبرز القدرة الفريدة لتقنيات المصفوفات على استخراج رؤى عميقة ومفيدة من البيانات المعقدة. خورشيد (2024) يشير إلى أهمية علم البيانات الضخمة في التطبيقات العملية، وهذه الدراسة تُجسد هذه الأهمية بطريقة ملموسة وقابلة للقياس.

نجاح النظام الهجين الذي يجمع بين تقنيات مختلفة في تحقيق أداء أفضل من الاعتماد على تقنية واحدة يُقدم درساً مهماً للممارسين في المجال. التحسن من معامل ارتباط 0.847 في النموذج الأساسي إلى 0.891 في النموذج الهجين الكامل يُثبت قيمة الدمج الذكي للتقنيات المختلفة. أبو رمان (2024) تؤكد على أهمية البيانات الضخمة في تحسين الأداء المنظمي، وهذا النهج المتكامل يُحقق تحسينات شاملة عبر مختلف جوانب الأداء.

التحديات التي واجهتها الدراسة، وخاصة في التعامل مع تناثر البيانات الشديد (99.33% قيم صفرية) والحاجة للتوازن بين التعقيد النموذجي والقابلية للتفسير، تُقدم دروساً قيمة للدراسات المستقبلية. الحلول المطورة للتعامل مع هذه التحديات، من استخدام تقنيات التخزين المتناثر إلى تطوير آليات التفسير التلقائي، يمكن تطبيقها في مشاريع مماثلة. مبدر (2024) يؤكد على أهمية قابلية إدارة تحليل البيانات الضخمة، وهذه الحلول تُساهم في تحسين هذه القابلية.

النتائج تفتح آفاقاً واسعة للتطبيق في مجالات متنوعة تتجاوز التجارة الإلكترونية، من الرعاية الصحية إلى التعليم والنقل والخدمات المالية. الإطار المنهجي المطور في هذه الدراسة يمكن تكييفه وتطبيقه في هذه المجالات مع التعديلات المناسبة لطبيعة كل مجال. اللهيبي (2023) تشير إلى أهمية إدارة المعرفة بناء على البيانات الضخمة، وهذا الإطار يُقدم أساساً قوياً لإدارة المعرفة في مختلف المجالات.

أخيراً، تُسلط هذه الدراسة الضوء على أهمية النهج التطبيقي في البحث العلمي، حيث أن الجمع بين الدقة النظرية والتطبيق العملي يُحقق قيمة حقيقية للمجتمع والاقتصاد. الأنصاري (2023) يؤكد على أهمية استخدام البيانات الضخمة في التطبيقات المختلفة، وهذه الدراسة تُقدم نموذجاً يُحتذى به في كيفية تحويل النظريات إلى حلول عملية. النجاح المحقق يُشجع على مزيد من الاستثمار في البحث التطبيقي ويفتح المجال أمام تطوير تقنيات أكثر تقدماً وفعالية في المستقبل.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

المراجع

قائمة المراجع باللغة العربية:

1. جاسم، عدى. (2012). تطبيقات لمصفوفة العوامل لحل نظام من معادلات Volterra التكاملية غير الخطية من النوع الثاني. مجلة أبحاث كلية التربية الأساسية، مج 12, ع 1 ، 773 - 785.
2. الدغاري، أكرم.(2017) . خوارزمية محسنة للعلامات المائية النصية في الصور الملونة (رسالة ماجستير غير منشورة). جامعة الشرق الأوسط، عمان.
3. محمد، الحسن. (2022). البيانات الضخمة: ماهيتها وأهميتها وعناصرها. المجلة العربية الدولية لإدارة المعرفة، مج1, ع2 ، 99 - 148.
4. يسري، حمزة. (2023). البيانات الضخمة: ثورة في عالم المعلوماتية وحل لتقصي قضايا الاستدامة. اقتصاديات - المجلة العربية للتنمية والعلوم الاقتصادية والمالية وعلوم التسيير، مج1, ع2 ، 50 - 66.
5. أحمد، الحسن شعبان. (2023). البيانات الضخمة وتحليلاتها: مراجعة علمية. المجلة الدولية للعلوم التربوية والإنسانية المعاصرة، مج2, ع2 ، 175 - 206.
6. عبدالحليم، أحمد. (2024). البيانات الضخمة وإجراءات المراجعة الخارجية: التحديات - الفرص: دراسة ميدانية. المجلة العلمية للدراسات والبحوث المالية والإدارية، مج16, ع4 ، 1 - 27.
7. عبدالحي، أميرة. (2022). البيانات الضخمة وتطبيقاتها وعلاقتها بالتنمية المستدامة. المجلة العربية للعلوم التربوية والنفسية، ع26 ، 161 - 180.
8. الدارودي، نهى. (2023). البيانات الضخمة ودورها في اتخاذ القرار. كتاب أعمال المؤتمر والمعرض السنوي السادس والعشرين: التقنيات الناشئة وتطبيقاتها في المكتبات ومؤسسات المعلومات، الكويت: جمعية المكتبات المتخصصة فرع الخليج العربي، 327 - 374.
9. الأنصاري، منذر. (2023). استخدام البيانات الضخمة في مجال الإحصاءات الرسمية. المجلة العربية للآداب والدراسات الانسانية، ع25 ، 401 - 422.
10. مبدر، غيث. (2024). قابلية إدارة تحليل البيانات الضخمة والأداء التنظيمي: الدور الوسيط للقابليات الديناميكية : دراسة على عينة من منظمات القطاع العام في محافظة ذي قار. الإدارة العامة، س64, ع3 ، 951 - 1030.
11. اللهيبي، ندى. (2023). إدارة المعرفة بناء على معالجة البيانات الضخمة: مراجعة لأدب الموضوع. مجلة دراسات المعلومات، ع26 ، 77 - 107.
12. خورشيد، معتز. (2024). علم البيانات الضخمة وتعلم الذكاء الاصطناعي. المجلة الدولية للتعليم بالإنترنت، 1 - 37.

(LOUJAS) **Copyright**: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

13. أبو رمان، إسراء. (2024). البيانات الضخمة وأثرها في البراعة المنظمية: دراسة حالة نظام "حكيم" مستشفى الحسين السلط الجديد. مجلة الزرقاء للبحوث والدراسات الإنسانية، مج24, ع1 ، 44 - 58.

قائمة المراجع باللغة الإنجليزية:

1. Luo, X., Zhou, M., Li, S., Hu, L., & Shang, M. (2020). Non-Negativity Constrained Missing Data Estimation for High-Dimensional and Sparse Matrices from Industrial Applications. *IEEE Transactions on Cybernetics*, 50, 1844-1855.
2. Luo, X., Zhou, M., Li, S., & Shang, M. (2018). An Inherently Nonnegative Latent Factor Model for High-Dimensional and Sparse Matrices from Industrial Applications. *IEEE Transactions on Industrial Informatics*, 14, 2011-2022.
3. Wu, D., & Luo, X. (2021). Robust Latent Factor Analysis for Precise Representation of High-Dimensional and Sparse Data. *IEEE/CAA Journal of Automatica Sinica*, 8, 796-805.
4. Huang, H. (2024). A Big Data-based Approach to Web User Behavior Analysis. *International Journal of Social Sciences and Public Administration*.
5. Jiang, D., Wang, Y., Lv, Z., Qi, S., & Singh, S. (2020). Big Data Analysis Based Network Behavior Insight of Cellular Networks for Industry 4.0 Applications. *IEEE Transactions on Industrial Informatics*, 16, 1310-1320.
6. He, J., & Yin, X. (2021). Internet User Behavior Analysis Based on Big Data. *2021 International Wireless Communications and Mobile Computing (IWCMC)*, 432-435.
7. Liu, Q., Wu, S., Wang, D., Li, Z., & Wang, L. (2016). Context-Aware Sequential Recommendation. *2016 IEEE 16th International Conference on Data Mining (ICDM)*, 1053-1058.
8. Gao, L. (2024). Design of Online Learning Behavior Analysis System Based on Big Data. *2024 Second International Conference on Data Science and Information System (ICDSIS)*, 1-5.
9. Hong, R. (2019). Research on Mobile User Behavior Mining Model Based on Big Data. *2019 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, 461-4613.
10. Bent, O., Dey, P., Weldemariam, K., & Mohania, M. (2017). Modeling user behavior data in systems of engagement. *Future Gener. Comput. Syst.*, 68, 456-464.
11. Liu, Z., & Song, T. (2024). Big Data Analysis and User Behavior Prediction of Social Networks Based on Artificial Neural Network. *Volume 31, Issue 3*.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **LOUJAS** and/or the editor(s). **LOUJAS** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

(LOUJAS) **Copyright**: © 2025 by the authors. Submitted for possible open access publication under the terms and .conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)