استخدام خوارزمية الجار الأقرب (Knn) في تحديد العوامل المؤثرة على خط الفقر الوطني

أ.م.د. جاسم الفارس م. نعمه عبد الله الفخري
جامعة الموصل / كلية الإدارة والاقتصاد

Using K-nearest neighbor (Knn) to determine influencing factors on the national poverty line

Assis.Prof.Dr. Jasim Al-Faris Lec .Nima Abdulla AL-Fakhry
College of Admin. &Eco. Mosul Uni.

تاريخ استلام البحث ٢٠١٦/٣/٢٧ تاريخ قبول النشر ٢٠١٦/٩/١٩

المستخلص:

يزداد التطور التقني للأجهزة والأنظمة الحاسوبية مع مرور الزمن، وكذلك شبكات الانترنيت والانترانيت، وأصبحت المؤسسات الحكومية وغير الحكومية في حاجة ملحة إلى التحول إلى عصر المعلومات وقواعد البيانات الكبيرة، لذلك تم البدء بإيجاد وسائل متطورة ومنها تقنيات الذكاء الاصطناعي. يعد التنقيب في البيانات إحدى تقنيات الذكاء الاصطناعي المتطورة فتم البدء بالعمل باستخدام تقنيات وأدوات التنقيب في البيانات لاستخراج المعرفة من مكامنها وتقليل الأخطاء واختصار الزمن.

يمثل البحث خطوة لتحديد بعض العوامل المؤثرة في خط الفقر الوطني في العراق وسبل معالجتها، وذلك من خلال الاستفادة من الإمكانيات الواسعة لتقنيات التنقيب في البيانات. تناول الجانب النظري من البحث مفهوم التنقيب في البيانات وأدواته وكذلك شرح خوارزمية الجار الأقرب (Knn) وتوضيح مفهوم الفقر، بينما تناول الجانب العملي تطبيق خوارزمية الجار الأقرب لمعرفة المؤشرات الأقرب لخط الفقر الوطني في العراق لسنة ٢٠١٢، وتم استخدام Matlab R2011a لتنفيذ الجانب العملي.

الكلمات المفتاحية: خوارزمية Knn، التنقيب في البيانات، لفقر، خط الفقر الوطني.

Abstract:

With over time the technical increases of sophistication devices and computer systems as well as networks of the Internet and Intranet, and the government institutions and non-governmental organizations became in an urgent need to shift to the information age and large databases, so it was starting to find an advanced methods such as artificial intelligence techniques. Data mining techniques are one of artificial intelligence techniques evolving to start using it's techniques and tools to extract knowledge from hideaways and reduce errors and save time.

The research identify some factors affecting the national poverty line in Iraq and ways to address them by taking the advantage of wide possibilities of data mining techniques. The theoretical side of this research talking about concept of data mining and tools, as well as explaining the classifier K-nearest neighbor (Knn) and clarify the concept of poverty, while the practical side of this research is application of an algorithm to determine the nearest neighbor indicators closes to the national poverty line in Iraq for 2012.

المقدمة:

تعد البيانات في عصر المعلومات والاتصالات والتقنيات الحديثة الحجر الأساس لاكتشاف المعرفة المخفية في كميات هائلة من البيانات، وأصبحت عملية استرجاع البيانات والحصول على المعلومات في حاجة إلى المزيد من عمليات البحث والتطوير لرفع مستوى الخدمة في قواعد البيانات الكبيرة.

التنقيب في البيانات هو الحل الأمثل لفتح الأبواب أمام اكتشاف المعرفة وإنتاج نماذج بيانية جديدة تقلل الجهد والوقت والكلفة، ويبحث في العلاقات بين مجاميع البيانات لإيجاد أنماط جديدة تفيد في تفسير المسائل وتوضيحها. وتعمل خوارزميات التصنيف على محاكاة العقل البشري الذي يعتمد على مستويات متباينة من الحقائق وأفرزت نماذج تصنيفية في شتى مجالات الحياة يمكن استخدامها في التحليل والتنبؤ واتخاذ القرارات.

تأتي فكرة البحث من منطلق الاستفادة من تقنيات التنقيب في البيانات في التعرف على مستوى الفقر في العراق من العلاقة بين المؤشرات المؤثرة في الفقر، والمؤشر الأقرب إلى خط الفقر الوطني بناءً على البيانات المتوفرة.

أهمية البحث:

تأتي أهمية البحث من:

١- أهمية دور التنقيب في البيانات في الحصول على المعلومات المهمة من البيانات الكثيرة والمتاحة لإيجاد الحلول المناسبة.

٢- دراسة الفقر في العراق باستخدام أدوات أكثر دقة بغية الحصول على أفضل الحلول.

أهداف البحث:

١- توضيح مفهوم التنقيب في البيانات.

٢- شرح خوارزمية الجار الأقرب.

٣- توضيح مفهوم الفقر والوقوف على بعض المؤشرات التي تؤثر في خط الفقر الوطني في العراق.

مشكلة البحث:

يشهد العراق الكثير من المشاكل والأزمات والفقر احد أهم تلك المشاكل ومن أجل التخفيف من هذه المشاكل وحلها لجأ البحث إلى تبني دور تقنيات المعلومات والاتصالات متمثلة بخوارزمية التصنيف (Knn) في الوقوف على أهم المؤشرات التي تؤثر على خط الفقر الوطني.

فرضية البحث:

تؤدي خوارزميات تصنيف البيانات دوراً أساسياً في تحديد مؤشرات الفقر التي تؤثر في خط الفقر الوطني في العراق. وبغية توضيح أهداف البحث وأهميته وإثبات فرضياته قسمنا البحث على المباحث الآتية:

المبحث الأول: التنقيب في البيانات

١- مفهوم التنقيب في البيانات:

قبل البدء بتوضيح مفهوم "**التنقيب في البيانات**" لابد من التمييز بين ثلاثة مصطلحات مهمة هي البيانات والمعلومات والمعرفة، إذ يقصد بالأولى تلك البيانات المقترنة بالعمليات الوظيفية للمنظمة، أي ما يتعلق بالأفراد والمواقع والصفقات، أما المعلومات فهي بيانات متماسكة تعطي رسالة معينة في حين ان المعرفة تضم العامل البشري فهي استنتاج يُرسم من المعلومات بعد ربطها مع معلومات أخرى ومقارنتها مع ما هو معروف سابقاً.

تجمع المنظمات مقدارا كبيرا "من البيانات، لكن الذي تريده حقا هو المعلومات، لذلك فإن التقنية الأحدث والأسرع للتعامل مع كمية البيانات، والإجابة عن أسئلة العمل، هي تنقيب البيانات التي تشبه عملية تنقيب الذهب فتنقيب الذهب عملية غربلة ما بين كميات من المعادن الخام لإيجاد كتلة من معدن ثمين ذي قيمة، وكذلك عملية تنقيب البيانات التي هي عملية غربلة ما بين كميات من البيانات لإيجاد المعلومات المفيدة لاتخاذ القرار (الدوري وأحمد،٢٠٠٧).

والتنقيب في البيانات هو أسلوب اكتشاف المعرفة الجديدة من كمية ضخمة من البيانات، والتي وجدت منذ فترة تاريخية لكنها مترابطة بسلسلة زمنية وبتحليل علائقي موضوعي، وباستخدام أساليب كانت موجودة لكنه لم تستثمر في هذا المجال مسبقاً مثل الذكاء الاصطناعي، وشجرة القرارات والنظم الخبيرة والعنقدة والخوارزميات الجينية والشبكات العصبية. مع تمثيل تلك المعارف المستكشفة بواحدة أو أكثر من تلك الأساليب وحسب طبيعة المشكلة. لذلك فإنَّ تقنية التنقيب تؤدي إلى إيجاد اتجاهات جديدة من البيانات وهما (سلمان،٢٠١٠):

١- **التنبوء الآلي المحوسب للصفات والاتجاهات:** تنقيب البيانات يعمل على إتمام عملية إيجاد المعلومات المتوقعة من قواعد البيانات الضخمة.

٢- **أتمتة استخراج الأنماط غير المعروفة سابقاً:** أدوات تنقيب البيانات يمكن أنْ تعرف شكلاً ذا نمط معين مختفي سابق.

يعد التنقيب عن البيانات تقنية تجمع بين الأساليب التقليدية في تحليل البيانات والخوارزميات المعقدة الخاصة بمعالجة مقادير ضخمة من البيانات، كم انها أتاحت فرصاً ممتعة من أجل استكشاف وتحليل أنواع جديدة من البيانات وتحليل أنواع قديمة بطرق جديدة.

فالتنقيب عن البيانات هو عملية استكشاف مؤتمتة لبيانات مفيدة في مصادر بيانات ضخمة ويعد التنقيب عن البيانات جزءاً مكملاً من اكتشاف المعرفة في قواعد البيانات وهي العملية الكلية التي يتم فيها تحويل البيانات الخام إلى معلومات مفيدة. (الطويل، ٢٠٠٩).

٢- الهدف من التنقيب في البيانات: (Nibest and et al, 2009)

- يهدف التنقيب في قواعد البيانات إلى استخلاص المعلومات المخبأة فيها، وهي تقنية حديثة، أصبحت مهمة في ظل التطور السريع وانتشار استخدام قواعد البيانات والمنافسة في السوق وغيرها.
- يوفر استخدامها القدرة على استكشاف أهم المعلومات في قواعد البيانات، والتركيز عليها في المؤسسات وأجهزة الأمن وفي غيرها من المجالات.
- تركز تقنيات التنقيب على بناء التنبؤات المستقبلية واستكشاف السلوك والاتجاهات، مما يسمح بتقدير القرارات الصحيحة واتخاذها في الوقت المناسب.
- تجيب تقنيات التنقيب على العديد من الأسئلة، وفي وقت قياسي، ولاسيما تلك النوعية من الأسئلة التي كانت من الصعب الإجابة عليها، إن لم يكن مستحيلاً، باستخدام تقنيات الإحصاء الكلاسيكية، والتي كانت إن وجدت فإنها تستغرق وقتاً طويلاً والعديد من إجراءات التحليل.

٣- أدوات التنقيب في البيانات:

هناك عدة أدوات تستخدم لأجل تنقيب البيانات، لذا فإن اختيار الأداة المناسبة يعتمد على طبيعة البيانات التي تتم دراستها فضلاً عن حجم البيانات، وهناك أنموذجان لتنقيب البيانات:

أ- **النموذج التنبؤي:** وهو النموذج الذي يستخدم النتائج المعروفة المستنبطة من البيانات المختلفة لأجل التنبؤ بقيم لاحقة للبيانات، ويتضمن هذا النموذج أهم أدوات التنقيب في البيانات:

١- **التصنيف:** يتم تحديد الأصناف قبل فحص البيانات ويتم تصنيف البيانات بناء على الخصائص المشتركة لها.

٢- **الانحدار:** هو تقنية تسمح بتحليل البيانات لوصف العلاقة بين متغيرين أو أكثر، إن الانحدار يفترض أنْ توضع البيانات بنوع معروف من الدوال ومن ثم يتم تحديد أفضل دالة للبيانات المعطاة.

٣- **تحليل السلاسل الزمنية:** إن مشاهدة البيانات عبر الزمن تنتج تحليلاً مفيدًا لأنه تتم مشاهدة سلوك البيانات عبر الزمن بشكل أساسي. وهذا يعني أن قيم الصفة المميزة للبيانات التي يتم فحصها تكون متغيرة عبر الزمن.

٤- **التنبؤ:** يعد التنبؤ من الأدوات التي تجذب الانتباه لأنها تتمكن من إعطاء مغزى التوقع الناجح في سياق العمل لذا فإنه يمكن النظر إلى العديد من تطبيقات تنقيب بيانات العالم الحقيقي كأنها تنبؤ بحالة بيانات مستقبلية معتمدة على بيانات سابقة وحالية.

ب- **النموذج الوصفي:** وهو النموذج الذي يُعَرِف الأنماط والعلاقات في البيانات، وعلى العكس من النموذج التنبؤي، فإن النموذج الوصفي يستخدم طريقة لاستكشاف خصائص البيانات التي تتم دراستها وليس للتنبؤ بخصائص جديدة، وهو من أكثر أدوات التنقيب في البيانات شيوعاً في هذا المجال، ويتضمن هذا النموذج الأدوات الآتية: (الياسري، ٢٠٠٥)

١- **قاعدة الارتباط:** قواعد الارتباط هي إحدى الواجهات الواعدة من قاعدة البيانات كأداة من أدوات اكتشاف المعرفة ولديها القدرة على تصفح كميات هائلة من البيانات، وهي تسمح بالتقاط كل القوانين الممكنة التي تشرح بعض الصفات الموجودة اعتمادا على وجود الصفات الأخرى. وبمعنى آخر هي قواعد ارتباطيه معينة بين مجموعة من البيانات في قاعدة البيانات، لقد ظهرت العديد من التعريفات لهذا المفهوم نذكر منها "التنقيب في مجموعة ضخمة من البيانات فضلا عن اكتشاف العلاقات" أو إنها "عملية استكشاف وتحليل كميات كبيرة من البيانات باستخدام أساليب آلية أو شبه آلية اعتماداً على اكتشاف نماذج وقواعد ذات مغزى".

٢- **العنقدة:** وهي عملية تقسيم البيانات إلى مجموعة من الأصناف اعتمادا على اشتراكها بالخواص المتشابهة وان العنقدة هي تقسيم غير موجه للبيانات، وهي عكس التصنيف الذي سيرد لاحقا، كما إنها تساعد المستفيد على فهم التركيب الطبيعي للمجموعات من البيانات.

٣- **التصنيف:** يستخدم التصنيف بشكل واسع في حل الكثير من المشكلات من خلال تحليل مجموعة من البيانات ووضعها على شكل أصناف أو أقسام يمكن استخدامها فيما بعد لتصنيف البيانات المستقبلية ،وهنا يكمن الفرق بين التصنيف والعنقدة. ويوجد العديد من الخوارزميات لتصنيف البيانات مثل الخوارزميات الإحصائية، وخوارزميات الشبكات العصبية، وخوارزميات الوراثة، وطريقة الجار الأقرب.... وغيرها.

٤- **التحليل التسلسلي:** في هذه الطريقة يتم البحث لاكتشاف نماذج تحدث بالتسلسل، إذ تكون المدخلات عبارة عن بيانات تشكل مجموعة متسلسلة وكل سلسلة من البيانات هي قائمة منظمة من العمليات أو المصطلحات وعندما تكون العملية عبارة عن مجموعات من المصطلحات لابد أن يحسب معها الوقت المصاحب لكل عملية.

٤- **خوارزمية الجار الأقرب (K_Nearest Neighbor KNN):**

هي إحدى خوارزميات التعليم الآلي والتي تعمل بمشرف (موّجه) وتعد خوارزمية الجار الأقرب من خوارزميات التصنيف التنبؤية والوصفية و لديها الإمكانيات لتوليد التقديرات المحلية للنقطة، وتكون هذه الطريقة سهلة في أغلب الأحيان عندما يكون حجم البيانات صغيراً، وتُصنِّف

البيانات لمعرفة الجار الأقرب بسهولة وفعالة عندما تكون قاعدة البيانات كبيرة جدا ولها القدرة على التعامل مع البيانات (المشوشة) أي الناقصة. وتعتمد الخوارزمية في عملها على قياس المسافة الاقليدية بين كل نقطة والنقطة الأقرب إليها وعندما تكون البيانات قريبة من بعضها تكون المسافة الاقليدية قليلة جدا بين كل نقطة والنقطة المجاورة لها ولكن كلما تباعدت قيم البيانات وتبعثرت أصبحت المسافات بين النقاط كبيرة ومنها جاء عنوان الخوارزمية، إذ يشير الحرف K إلى الحالات التي سيتم تصنيفها بناء على المسافات بينها (أي بين الجيران)، وتُحسب المسافة من المعادلة الآتية: (الطائي وهرمز، ٢٠١١) (Imandoust, and Bolandraftar, 2013)

$$d(i,j) = \sqrt[2]{\left|x_i - x_j\right|^2} \quad \text{..... (1)}$$

حيث إنَّ:

d: المسافة بين أي نقطتين.

x_i, x_j: نقاط البيانات في فضاء البحث.

ولهذه الخوارزمية محددات مثلما لها مزايا، ومن أهم محدداتها انها تحتاج إلى وقت طويل عند العمل على البيانات الكبيرة، وكذلك فهي حساسة عند التعامل مع مجاميع التدريب ذات الخواص غير المترابطة؛ لأنَّ هذا يُضعف التصنيف ولهذا يجب اختيار مجاميع التدريب ذات الخواص المتقاربة لتقوية النموذج (Imandoust, and Bolandraftar, 2013).

- **مراحل عمل الخوارزمية:** (خير الدين وعبد ، ٢٠١٢: ٤٢٩)

- حدد عدد الجيران الأقرب ولتكن K.
- حساب المسافة الأقليدية بين السجل المستكشف وأقرب جار من المعادلة (١).
- تُرتب المسافات بإعطاء الرتب لها من أصغر مسافة إلى أعلى مسافة، ثم حدد الجيران الأقرب بالاستناد إلى مسافة حدّ أدنى k-th.

المبحث الثاني: الفقر

١- مفهوم الفقر:

الفقر مشكلة اقتصادية عالمية ذات أبعاد وامتدادات اجتماعية متعددة، وهي ظاهرة تكادُ لا يخلو منها أي مجتمع، مع التفاوت في حجمها والآثار المترتبة عليها، هذا وأشارت التقديرات إلى أنَّ حوالي خمس العالم يمكن تصنيفهم على أنَّهم فقراء تنقصهم الحدود الدنيا من فرص العيش الكريم. إنَّ مواجهة الفقر والتخفيف منه تمر عبر مراحل متعددة تبدأ بقياس الفقر متضمنة تعريف من هم الفقراء، فمفهوم الفقر يكاد يتسم بالتفاوت تبعا لتفاوت المجتمعات من حيث المستوى الاجتماعي والاقتصادي، فما هو فقرٌ في دولة معينة قد لا يكون فقرا في دولة أخرى.

يختلف مفهوم الفقر باختلاف البلدان والثقافات والأزمنة ولا يوجد إتفاق دولي حول تعريف الفقر نظراً لتداخل العوامل الاقتصادية والاجتماعية والسياسية التي تشكل ذلك التعريف وتؤثر عليه، وعليه فهناك اتفاق حول مفهوم الفقر على أنه حالة من الحرمان المادي الذي يترجم بانخفاض

استهلاك الغذاء، كماً ونوعاً، وتدني الوضع الصحي والمستوى التعليمي والوضع السكني، والحرمان من السلع المعمرة والأصول المادية الأخرى، وفقدان الضمانات لمواجهة الحالات الصعبة كالمرض والإعاقة والبطالة وغيرها. وبمفهوم مبسط للفقر يعيش الفرد أو الأسرة ضمن إطار الفقر إذا كان الدخل المتأتي له غير كافٍ للحصول على أدنى مستوى من الضروريات للمحافظة على نشاطات حياته وحيويتها (الرفاعي، ٢٠٠٧).

٢- أسباب تفشي الفقر:

من أهم الأسباب التي أدت إلى تفشي ظاهرة الفقر ما يأتي:

أ) **حجم الأسرة:** إن حجم الأسرة يعد أيضا من مسببات الفقر، إذ يؤدي كبر حجم الأسرة و ارتفاع معدلات الإعالة إلى زيادة الأعباء على نفقات الأسرة وبالتالي مواجهة حالة العجز عن توفير كل متطلباتها ذات الحجم الكبير، ومع استمرار حالة العجز وتناميها ينتج عنها الفقر بأتم معناه.

ب) **التضخم:** إن التضخم الذي يعرف بأنه الارتفاع العام في أسعار السلع والخدمات معبراً عنها بالنقود يؤدي إلى انخفاض القوة الشرائية للنقود و بالتالي تتأثر الدخول الحقيقية للأسر و تصل إلى حالة العجز عن اقتناء كل المتطلبات التي تحتاجها و تصبح ضمن تعداد الفقراء بغض النظر عن درجة الفقر.

ج) **برامج التصحيح الهيكلي:** تُعَدُّ برامج التعديل أو التصحيح الهيكلي واحدة من أهم الأسباب التي أدت إلى تنامي الفقر و ازدياد معدلاته، لاسيما في العالم النامي، فقد كان الكثير من الدول النامية قد عرفت تدهوراً شديداً في الظروف الاجتماعية مع تزايد سوء التغذية، وبطء التحسينات في مجال الصحة أو تراجعها، وانخفاض مستوى التعليم... وما إلى ذلك.

د) **النزاعات الداخلية و الخارجية:** كالحروب مثلا تساهم الحروب على سبيل المثال في الاستقرار و ما ينتج عنه من ضياع فرص العمل و ضياع الممتلكات و غيرها و بالتالي السير نحو الفقر.

هـ) **سوء توزيع الدخل والثروات:** إن غياب التوزيع العادل للدخل القومي والثروات يؤدي إلى غناء البعض و إفقار البعض الآخر.

و) **تدني المستوى التعليمي:** أصبح من المعترف به لدى الجميع اليوم، أن المشكلات والقضايا التربوية والثقافية (محو الأمية وإنشاء الأنظمة الوطنية للتعليم وترقيتها وبلوغ وضع جديد كيفياً للثقافة) لبلد ما من البلدان تؤثر في مجال التحولات الاقتصادية والاجتماعية وتشارك بها. وليس من قبيل الصدفة أن نلاحظ العلاقة الجدلية بين الجوع والأمية والتخلف الاقتصادي. كما أن انتكاسات الأمة العربية الحضارية والتاريخية تعود في أغلبها إن لم تكن كلها إلى حالة التخلف والجهل والتبعية التي تقف شاهد عيان على فشل تلك المشاريع الفكرية العربية.

ي) **البطالة:** على الرغم من نُدرة الإحصاءات الحديثة والمنظمة عن معدلات البطالة في الدول العربية، وعدم تطابق أو اتفاق ما هو متاح منها، إلا أنها تعبر بشكل عام عن ضخامة ظاهرة البطالة، بين

الشباب خاصة. وتُشير بيانات منظمة العمل العربية إلى أن ظاهرة بطالة حملة الشهادات التعليمية قد استفحلت في العديد من الدول العربية، وهذا يدل أن معدلات البطالة في أوساط الشرائح الشابة من المجتمع مرتفعة، وتتجاوز المقاييس والمعدلات العالمية. وأن تفاقم هذه الظاهرة في المجتمع يُفرز جملة من الآثار والانعكاسات الاجتماعية.فهي قد تؤدي تدريجيا إلى سلسلة من الحالات المختلفة، أي من البطالة إلى الإقصاء، ومن الإقصاء إلى التهميش، ومن التهميش إلى الجنوح، كما أنها تزيد من حدة الفوارق الاجتماعية والشعور بعدم المساواة بين أفراد المجتمع.

ز) **التوزيع غير العادل للثروات الوطنية:** "العالم اليوم أصبح جزيرة أغنياء تحيط بها بحار من الفقراء" هكذا وصف الرئيس الجنوب أفريقي "مبيكي" معضلة الفقر التي تزداد يوماً بعد يوم رغم التقدم الذي أحرزته البشرية في شتى المجالات، ورغم جني الكثير من خيرات الكوكب التي يُجمع الخبراء على أنها كافية لتقديم الرفاهية للستة مليارات من البشر الذين يعيشون فوقه لو تم توزيعها بالحد الأدنى من العدالة (محمد وآخرون، ٢٠١١) .

– دور تقنيات المعلومات والاتصالات في مكافحة الفقر:

تعد تقنيات المعلومات والاتصالات أداة تنموية ينبغي أن يكون الهدف من استعمالها ونشرها هو تحقيق فوائد في كل جوانب حياتنا اليومية، وذلك من خلال تطبيقاتها على صعيد العمليات والخدمات الحكومية، والرعاية الصحية، والمعلومات الصحية، والتعليم، والتدريب والعمل، وتوفير فرص العمل والأعمال التجارية، والزراعة، والنقل، وحماية البيئة، وإدارة الموارد الطبيعية ومنع الكوارث، والثقافة، واستئصال الفقر وغيرها، ويتلخص دور تقنيات المعلومات والاتصالات في مكافحة الفقر على الفعاليات الاقتصادية والاجتماعية المختلفة كافة وبالتالي على الاقتصاد العام. وهذا يؤكد بان لقطاع تكنولوجيا المعلومات والاتصالات دوراً تنموياً على جميع الأصعدة وخاصة في تعزيز التنمية المستدامة وعلى مكونات الفقر الرئيسية من تعليم، وصحة، ودخل، وتشغيل. ويتلخص الدور في تحسين مستوى المعيشة ومحاربة الفقر في الدول النامية، من حيث خلق وظائف جديدة وتحسين مستوى أداء الأفراد، وأسلوب العمل، وتحسين الخدمات المقدمة للأفراد بما فيهم الفئات المهمشة، وتخفيض التكلفة (المجلس الاقتصادي الفلسطيني للتنمية والاعمار،٢٠٠٨)

المبحث الثالث: الجانب العملي

تمهيد:

العراق من الدول الغنية بالكثير من الثروات الطبيعية التي من المفترض انها تدر عليها دخلاً كبيراً، ولكن سنواتٍ طويلة من الحروب وعدم الاستقرار في الجوانب السياسية والاقتصادية والاجتماعية أدت إلى انتشار الفقر وتفشيه في البلاد، ولأجل الوقوف على أهم المؤشرات المؤثرة على خط الفقر الوطني فقد سعى البحث إلى تحليل تلك المؤشرات مستخدما خوارزمية الجار

الأقرب (Knn) لمعرفة وتحديد أي من تلك المؤشرات هي الأقرب إلى مؤشر خط الفقر. وفيما يلي مراحل الجانب العملي:

١- البيانات المعتمدة في البحث:

اعتمد الباحثان على بعض مؤشرات الفقر في العراق للعام ٢٠١٢ والمأخوذة من البنك الدولي وتقارير التنمية البشرية لدى الأمم المتحدة وكما يوضح الجدول (١):

جدول رقم (١)

بعض مؤشرات الفقر في العراق/ ٢٠١٢

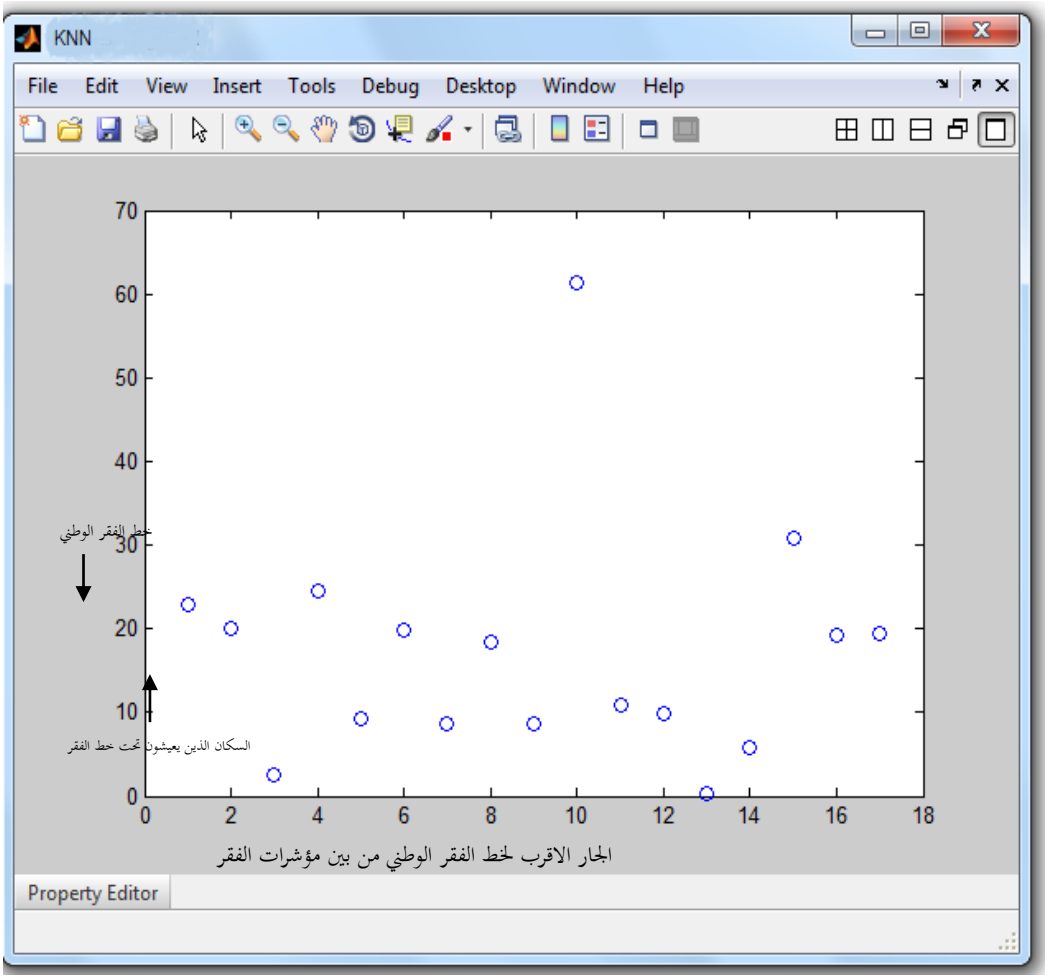
المؤشر	النسبة المئوية
خط الفقر الوطني	22.9
السكان الذين يعيشون تحت خط الفقر	2.8
مستوى المعيشة	20.4
التعليم (الحرمان في الفقر العام)	47.5
الصحة (الحرمان في الفقر العام)	32.1
السكان الذين يعيشون في فقر مدقع	3.1
السكان المعرضون لخط الفقر	14.3
شدة الحرمان (السكان في فقر متعدد الابعاد)	41.3
مجموع السكان (السكان في فقر متعدد الابعاد)	14.2
الإعانة الإجمالية (١٥ سنة-٦٤ سنة)/١٠٠ شخص	84.3
البطالة من إجمالي القوى العاملة	12.1
البطالة من الذكور/ إجمالي القوى العاملة	13.1
البطالة من الإناث/ إجمالي القوى العاملة	22.5
البطالة عند الشباب (١٥-٢٤) سنة	28.6

البطالة عند الشابات (١٥-٢٤) سنة	53.7
الإنفاق على الرعاية الصحية	3.6
نصيب الفرد من الناتج المحلي الإجمالي	3.557

المصدر : www.albankaldawli.org و تقارير التنمية البشرية للأمم المتحدة/٢٠١٢.

٢- تنفيذ خوارزمية الجار الاقرب (Knn):

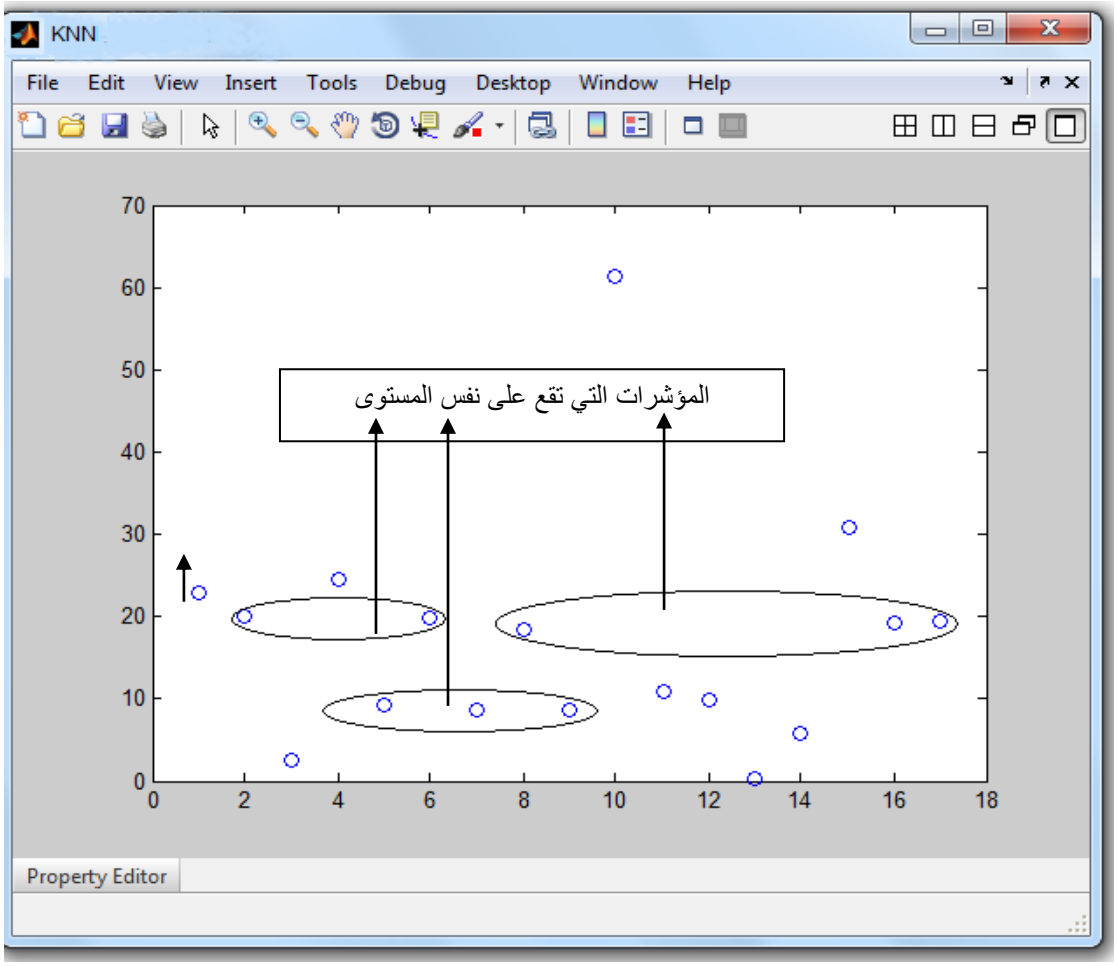
تمثل خوارزمية الجار الأقرب المسافة الاقليدية بين إحدى نقاط البيانات والنقاط الأخرى الموجودة في متجه التدريب والذي يمثل في هذا البحث مجموعة المؤشرات في الجدول (١)، وتم احتساب المسافة الاقليدية بين مؤشر خط الفقر الوطني والمؤشرات الأخرى من المعادلة (١)، والشكل رقم (١) يوضح المسافة بين مؤشرات الفقر ومؤشر خط الفقر الوطني:



شكل رقم (١)

نتائج تنفيذ خوارزمية الجار الأقرب (Knn)

نلاحظ من الشكل السابق أن اقرب جار لمؤشر خط الفقر الوطني هو المؤشر الذي يمثل السكان الذين يعيشون تحت خط الفقر، وكان أبعد مؤشر هو مؤشر الإعانة الإجمالية، ولوحظ من الشكل أيضا ان هناك مؤشرات تقع على المستوى نفسه من مؤشر (السكان الذين يعيشون تحت خط الفقر) ولا يمكن إهمال المؤشرات التي تكون متقاربة من بعضها، والشكل (٢) يوضح المؤشرات المتقاربة من بعضها، أي انها تعنقدت مع بعضها:



شكل رقم (2)

المؤشرات المتعنقدة مع بعضها

ولأجل التخلص من حالة ضبابية المسافة بين المؤشرات فقد اقترح الباحثان دالة انتماء (دالة عضوية) لتحديد درجة الانتماء لكل مؤشر وتوضيح المؤشرات المتقاربة مع بعضها متمثلة بالمعادلة (2)، والجدول (٢) والشكل رقم (٣) يوضحان تنفيذ دالة العضوية على المؤشرات:

$$\mu(x)=\begin{cases}1 & \text{if } x=22.9\\ \dfrac{\text{distance}\,(x_i-x_j)}{\overline{X}}-1 & \\ 0 & \text{other wise}\end{cases} \qquad \ldots\ldots (2)$$

حيث ان:

X: المعدل.

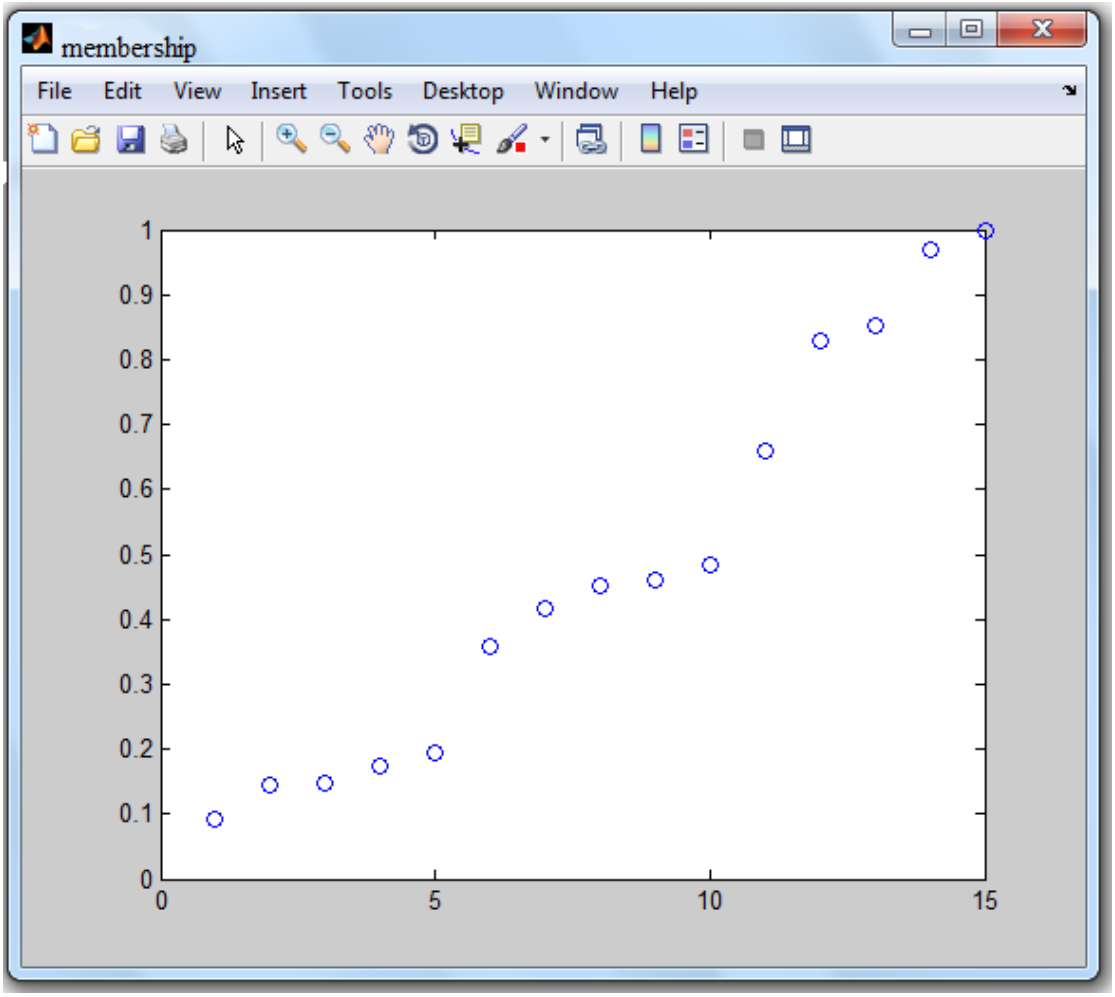
distance (x_i,x_j): المسافة الاقليدية بين النقطتين.

جدول رقم (٢)

دالة الانتماء (العضوية) لمؤشرات الفقر

المؤشرات	البيانات%	دالة الانتماء(العضوية)
خط الفقر الوطني	22.9	1
البطالة من الإناث/ إجمالي القوى العاملة	22.5	0.976247031
مستوى المعيشة	20.4	0.851543943
البطالة عند الشابات (١٥-٢٤) سنة	53.7	0.828978622
البطالة عند الشباب (١٥-٢٤) سنة	28.6	0.66152019
السكان المعرضون لخط الفقر	14.3	0.489311164
مجموع السكان (السكان في فقر متعدد الابعاد)	14.2	0.483372922
التعليم (الحرمان في الفقر العام)	47.5	0.460807601
الصحة (الحرمان في الفقر العام)	32.1	0.45368171
البطالة من الذكور/ إجمالي القوى العاملة	13.1	0.418052257
البطالة من إجمالي القوى العاملة	12.1	0.358669834
السكان الذين يعيشون تحت خط الفقر	2.8	0.193586698
السكان الذين يعيشون في فقر مدقع	3.1	0.175771971

نصيب الفرد من الناتج المحلي الإجمالي	3.557	0.148634204
الإنفاق على الرعاية الصحية	3.6	0.14608076
شدة الحرمان (السكان في فقر متعدد الابعاد)	41.3	0.09263658



شكل رقم (٣)

دالة العضوية لمؤشرات الفقر

نلاحظ من الشكل رقم (٣) ان المؤشرات التي تقع على مستويات متقاربة من بعضها قد ترتبت بأربعة مستويات، وكان أقرب مؤشر لخط الفقر الوطني هو نسبة البطالة من الإناث وتلاها مؤشرا (مستوى المعيشة والبطالة عند الشابات) وجاءت مؤشرات (السكان المعرضون لخط الفقر والتعليم وعدد السكان و الصحة) في مستوى واحد تقريباً، وأخيراً وقعت مؤشرات (السكان الذين يعيشون تحت خط الفقر والسكان الذين يعيشون في فقر مدقع ونصيب الفرد من الناتج المحلي الإجمالي ونصيب الفرد من الإنفاق على الرعاية الصحية وشدة الحرمان) في مستوى واحد.

الاستنتاجات والمقترحات

أولا: الاستنتاجات

١- إن معالجة البيانات والمعلومات وتصنيفها تكون خالية من الدقة، إذا لم يتم تنظيمها وتوصيفها بشكل جيد.

٢- دلت نتائج التطبيق الفعلي على أن استخدام خوارزمية الجار الأقرب ودالة الانتماء المقترحة للتصنيف اظهر نتائج دقيقة و واضحة.

٣- خوارزمية الجار الأقرب هي الأنسب في تحديد المؤشرات القارب إلى مؤشر خط الفقر؛ لأنها تعتمد حساب المسافة الاقليدية بين كل مؤشر من المؤشرات وخط الفقر وبالتالي اتخاذ القرار المناسب بشأنها.

٤- إن أقرب مؤشر لخط الفقر هو مؤشر السكان الذين يعيشون تحت خط الفقر وهذا أمر طبيعي حسب المفهوم الاقتصادي عن خط الفقر.

٥- إن ابعد مؤشر هو مؤشر الإعانة وذلك لان غالبية المعونات لا تذهب إلى مستحقيها وبالتالي فإن أثرها يكون بعيد بسبب الأوضاع غير المستقرة للبلد.

٦- أثرت باقي المؤشرات بنسب متقاربة على خط الفقر وكما موضح في الاشكال المرفقة بالبحث.

ثانياً: المقترحات

١- الاهتمام بتطبيق الإصدارات الأحدث من هذه الخوارزمية مثل Weighted Knn و Fuzzy Knn ومؤشرات أخرى للإحاطة بمحددات الفقر في العراق والوطن العربي.

٢- تهيئة الظروف الملائمة والاحتياجات المادية والمعنوية لتقليص فجوة الفقر في العراق.

٣- إعادة توزيع الإعانات على المحتاجين الأساسيين ووضع خطط اقتصادية مستقبلية لتقليل نسبة الفقر في العراق وتوزيع الإعانات توزيعاً عادلاً.

المصادر

أولا: العربية

١- البحوث والدوريات:

- الدوري، زكريا مطلك و أحمد، داليا عبد الحسين (٢٠٠٧)، "**دور التنقيب في البيانات Data Mining في زيادة أداء المنظمة–دراسة تحليلية في المصرف الصناعي**"، مجلة العلوم الاقتصادية والإدارية، المجلد ١٣، العدد ٤٨، ص ٤٣، كلية الإدارة والاقتصاد– جامعة بغداد، بغداد، العراق.
- الرفاعي، محمد عبد الله (٢٠٠٧)، "**معوقات بيانات قياس الفقر**"، المؤتمر الإحصائي العربي الأول"، عمان– الأردن، ص ٤٦٩-٤٧٠.
- الطائي، فاضل عباس و هرمز، ساندي (٢٠١١)، "**التنبؤ بالسلسلة الزمنية باستخدام طريقة الجار الأقرب المضبب مع التطبيق**"، المجلة العراقية للعلوم الاحصائية، ص١٨١–١٩٦.
- المجلس الاقتصادي الفلسطيني للتنمية والاعمار– بكدار (٢٠٠٨)،"**دور تكنولوجيا المعلومات والاتصالات التنموي ومكافحة الفقر في فلسطين**"، ص ١٧.
- الياسري، أروى عيسى (٢٠٠٥)، "اتجاه جديد في استرجاع المعلومات"، مقالة، مجلة المعلوماتية، العدد ١٦، ص ٧-٨، www.informatics.gov.sa.
- خير الدين محمد، انهار و عبد داود، سهير (٢٠١٢)، "**Development of important patterns in student database using statistical classifier**", مجلة جامعة الانبار للعلوم الاقتصادية والإدارية، المجلد ٤ العدد ٩، ص ٤٢٩.
- سلمان، عبد الستار شاكر (٢٠١٠)، "**التنقيب والعادة هندسة قواعد البيانات**"، **المؤتمر السادس لجمعية المكتبات والمعلومات السعودية: (البيئة المعلومات الآمنة: المفاهيم والتشريعات والتطبيقات)** ، الرياض٢١-٢٢ ربيع الثاني/١٤٣١ هـ ٦-٧ ٢٠١٠ م، ص ١١.
- محمد، عجيلة و مصطفى، بن نوي و سيد علي، بلحمدي (٢٠١١)، "**إستراتيجية معالجة الفقر في ظل العولمة–حالة الجزائر**"، المنتدى العربي للعلوم الاجتماعية والإنسانية، ص ٤-٩.

٢– الكتب:

- الطويل، هالة (٢٠٠٩) ، "**المرجع التعليمي في التنقيب عن البيانات**"، شعاع للنشر والتوزيع، ط١، ص ٨-٩، سوريا، حلب.

ثانياً: الأجنبية

- Imandoust, Sadegh Bafandeh and Bolandraftar, Mohammad (2013), "**Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background",** Journal of Engineering Research and Applications, Vol. 3, Issue 5, PP.605-610.

- Nisbet, & Elder, John & Miner, Gary (2009) "**Handbook of Statistical Analysis & Data Mining Applications**", Academic Press /Elsevier, ISBN 9780123747655, P.12.