1.1 المقدمة

التنقيب في البيانات هو عملية تحليل كمية كبيرة من البيانات لإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة لصاحب البيانات، كما يمكن من الوصول إلي المعلومات التي تساعد في صناعة القرار [1]. إن من اهم الاهداف للتنقيب في البيانات هو التنقيب الوصفي وهو إنتاج معلومات جديدة بناء علي المعلومات الموجودة داخل البيانات المستخدمة في عملية التنقيب ومن العمليات الأساسية التي تعتبر ضمن عمليات او مهام التنقيب في البيانات هي عملية العنقدة.

العنقدة من التقنيات الشائعة والمفيدة في اكتشاف المجاميع المتعددة وتعريف الأنماط والتوزيعات المهمة في البيانات الخام. إن الفكرة الأساسية للعنقدة تلخص بتجزئة البيانات إلى عناقيد اذ إن نقاط البيانات للعنقود الواحد تتشابه أكثر مع بعضها البعض مقارنة مع تلك النقاط في العناقيد الأخرى.[2]

قام المعهد العالي لعلوم الزكاة بعمل مسح شامل للفقراء ويقصد بالمسح الحصر الشامل للفقراء في السودان وتم من خلاله تقسيم الفقراء الي درجات سماها درجات عناصر الحياة الطيبة وعرفها علي انها مجموعة من المكونات التي يجب علي الانسان استهلاكها او تعاطيها ليعيش بحرية وكرامة وتشمل الصحة، المياه، التعليم، المأكل، الملبس والمأوي. وقد قام بإعطاء كل مكون منها درجة تمثل أهميتها في عناصر الحياة الطيبة . وبحسب التوليفة التي يتعاطاها الإنسان ومستواها من هذه المكونات يُعد غنيا أو فقيرا .[3]

1.2 مشكلة البحث

السؤال الذي يناقشه هذا البحث تكمن في معرفة هل تصنيف فئات الفقر الذي تتبعه الامانة العامة لديوان الزكاة مثالي ام ان هنالك تصنيف آخر أكثر ملائمة.

1.3 أهداف البحث

يهدف هذا البحث الي دراسة بعض طرق العنقدة واستخدامها لتصنيف بيانات ديوان الزكاة.

1.4 منهجية البحث

في هذه الدراسة سيتم اتباع منهجية اكتشاف المعرفة في قاعدة البيانات (KDD) [2] على قاعدة بيانات الحصر الشامل للاسر الفقيرة للعام 2011 وهي بيانات تخص حالات الفقر علي نطاق السودان تم جمعها بواسطة المعهد العالي لعلوم الزكاة. مستخدمين واحدة من طرق العنقدة وهي خوارزمية k-means وذلك بأداة التطبيق Weka .

مراجعة الادبيات

2.1 ماذا يعني التنقيب عن البيانات

أدى انتشار الواسع لتقنية المعلومات وسهولة إتاحتها إلى تضخم حجم المعلومات بصورة استباقية لم يشهدها التاريخ من قبل، مما جعل من قضية البيانات الضخمة على الإنترنت مثار للجدل، من حيث جدوى وجودها بهذه الصورة العشوائية. وعندما نتحدث عن البيانات الضخمة، فإننا نتحدث عن كميات لايمكن تخيلها من البيانات المتعددة الآنواع. معلومات في كل مكان وفي موضوعات شتي. وهنا نتساءل ما أهمية هذه البيانات في ظل وجود حقيقة تشير إلى أن المعلومات المنظمة من هذه البيانات لا تشكل إلاجزءً ضئيلا يصل الي 10 % مقارنة بالمعلومات غير المنظمة والتي تشكل الباقي. ذلك ادي الي ازدياد الحاجة الي تطوير ادوات تمتاز بالقوة لتحليل البيانات واستخراج المعلومات والمعارف منها. فالاساليب التقليدية والاحصائية لا تستطيع ان تتعامل مع هذا الكم الهائل لذا تستخدم ادوات ذكية لمعالجة هذه البيانات.

ومن هنا ظهر ما يسمي بتنقيب البيانات Data Mining كتقنية تهدف الي استنتاج المعرفة من كميات هائلة من البيانات ، تعتمد علي الخوارزميات الرياضية والتي تعتبر اساس التنقيب عن البيانات وهي مستمدة من العديد من العلوم مثل علم الاحصاء والرياضيات والمنطق وعلم التعلم ،والذكاء الاصطناعي والنظم الخبير، وعلم التعرف علي الانماط ، وعلم الاله. وغيرها من العلوم التي تعتبر من العلوم الذكية وغير التقليدية.

ظهر التنقيب في البيانات (Data Mining) في اواخر الثمانيات وأثبت وجوده كأحد الحلول الناجحة لتحليل كميات ضخمة من البيانات، وذلك بتحويلها من مجرد معلومات متراكمة وغير مفهومة (بيانات) الي معلومات قيمة يمكن استغلالها والاستفادة منها بعد ذلك.

وقد اجتذبت مرحلة التنقيب في البيانات الكثير من الاهتمام في الآوساط البحثية على مدى العقد الماضي، في محاولة لتطوير خوارزميات قابلة للتوسع والتكيف مع كميات متزايدة من البيانات في البحث عن انماط معرفية ذات معني. وقد نمت حزم من الخوارزميات والبرمجيات وبشكل كبير خلال العقد الماضي، الي حد ان التوسع قد جعل من الصعب علي العاملين في هذا الحقل تتبع التقنيات المتاحة لحل مهمة معينة. [2]

2.2 مراحل اكتشاف المعرفة

اكتشاف المعرفة في قواعد البيانات Knowledge discovery in Database(KDD) ليس بالعملية السهلة والتي قد يعتقد البعض انها تتوقف عند تجميع البيانات وإدارتها، بل نراها تمتد الي التحليل والتوقع والتنبؤ بما سيحدث مستقبلاً.

التنقيب في البيانات يشكل جزءاً من اكتشاف المعرفة Knowledge discovery، وهذه العملية هي الاكثر شمولاً .

2.3 تتضمن عملية اكتشاف المعرفة:

- اكتشاف البيانات data discovery : وهي مرحلة جمع البيانات وتشمل كشف وتحديد وتوصيف البيانات المتاحة.
- تصفية البيانات وتنقيتها Data cleaning : ويتم في هذه المرحلة ازالة البيانات المزعجة Noise والتي لا اهمية لها، كما يتم حذف البيانات المتضاربة والبيانات الغير متناسقة.
- تكامل البيانات Data Integration : يتم في هذه المرحلة تجميع البيانات المتشابهة وذات الصلة من مصادر البيانات المتعددة ودمجها معاً.
- اختيار البيانات Data selection : في هذه المرحلة يتم تحديد واسترجاع البيانات الملائمة من مجموعة البيانات.
- تحول البيانات Data Transformation : في هذه المرحلة يتم تحويل البيانات الي نماذج مخصصة ملائمة لاجراءات البحث والاسترجاع بواسطة خلاصة الانجاز او عمليات التجميع.
- التنقيب عن البيانات Data Mining: أي استخدام طرق ذكية تطبق لاستخلاص انماط البيانات واستخراج نماذج مفيدة قدر الامكان.
- تقييم الانماط Pattern evaluation: يتم في هذه المرحلة تحديد الانماط المهمة حقا والتي تمثل قاعدة المعرفة لاستخدام بعض المقاييس المهمة.
- تمثيل المعرفة Knowledge presentation : وهي المرحلة الاخيرة من مراحل اكتشاف المعرفة في قواعد البيانات وهي المرحلة التي يراها المستفيد، هذه المرحلة الاساسية تستخدم الاسلوب المرئي لمساعدة المستفيد في فهم وتفسير نتائج استخراج البيانات.

وبالتالي التنقيب في البيانات هو خطوة اساسية لتطبيق اساليب ذكية بهدف الكشف عن أنماط البيانات المثيرة للاهتمام والمخبأة في مجموعات البيانات الكبيرة , ومع ذلك في بعض المنظمات نحد أن مصطلح التنقيب عن البيانات (Data Mining) اصبح اكثر شعبية للاشارة الي العملية التي يتم فيها اكتشاف المعرفة(Knowledge Discovery) برمتها.

وهناك جانب مهم جداً ، يجب النظر اليه بعين الاعتبار وهو أن هناك انماط جديدة قد تبرز، عدة ما تكون غير معروفة من قبل. وبالتالي يجب ان تكون أدوات التنقيب عن البيانات قادرة علي البحث عن أنواع مختلفة من الانماط، باشكال متوازية لزيادة كفاءة التنقيب عن البيانات. كما يجب ايضاً ان يتم الكشف عن الانماط في الاجزاء الصغيرة والفرعية، لأن بعض الأنماط قد لاتصلح لكافة البيانات في مجموعة البيانات التي تم تحليلها.[2]

2.4 التنقيب في البيانات

يعرف التنقيب في البيانات علي انه عملية تحليل كمية بيانات (عادة ما تكون كمية كبيرة) لايجاد علاقة منطقية تخلص البيانات بطريقة جديدة تكون مفهومة ومفيدة لصاحب البيانات. يطلق اسم (نماذج Models) علي العلاقات والبيانات الملخصة التي يتم الحصول عليها من التنقيب في البيانات. يتعامل تنقيب البيانات عادة مع بيانات يكون قد تم الحصول عليها بغرض غير غرض التنقيب في البيانات(مثلاً قاعدة بيانات التعاملات في مصرف ما) مما يعني ان طريقة التنقيب في البيانات لا توثر مطلقاً علي طريقة تجميع البيانات ذاتها. هذه هي احد النواحي التي تختلف فيها التنقيب في البيانات علي الاحصاء ، ولهذا يشار الي عملية التنقيب في البيانات علي انها عملية إحصائية ثانوية. يشير التعريف أيضاً الي ان كمية البيانات تكون عادة كبيرة، اما في حال كون كمية البيانات صغيرة فيفضل استخدام الطرق الاحصائية العادية في تحليلها.

عند التعامل مع حجم كبير من البيانات تظهر مسائل جديدة مثل كيفية تحديد النقاط المميزة في البيانات، وكيفية تحليل البيانات في فترة زمنية معقولة وكيفية قرار ما اذا كانت أي علاقة ظاهرية تعكس حقيقة في طبيعة البيانات. عادة يتم التنقيب في بيانات تكون جزءاً من البيانات حيث يكون الغرض عادة تعميم النتائج علي كامل البيانات (مثلاً تحليل البيانات الحالية لمستهلكي منتج ما بغرض توقع طلبات المستهلكين المستقبلية). من احد اهدف تنقيب البيانات ايضاً هو اختزال كميات البيانات الكبيرة او ضغطها بحيث تعبر بشكل بسيط عن كامل البيانات بدون تعميم.

2.5 الأدوات الأساسية لتنقيب البيانات Basic Tools of Data Mining

هنالك عدة ادوات تستخدم لاجل تنقيب البيانات لذا فان اختيار الادوات المناسبة يعتمد علي طبيعة البيانات التي تتم دراستها فضلاً عن حجم البيانات، وهناك نموذجان لتنقيب البيانات.

اولاً: النموذج التنبؤي PredictiveModel : وهو النموذج الذي يستخدم النتائج المعروفة المستبطة من البيانات المختلفة لاجل التنبؤ بقيم لاحقة للبيانات ، ويتضمن هذا النموذج اشهر ادوات تنقيب البيانات.

- التصنيف Classification : يتم تحديد الاصناف قبل فحص البيانات ويتم تصنيف البيانات بناءً علي الخصائص المشتركة لها.
- الانحدار Regression : وهو تقنية تسمح بتحليل البيانات لوصف العلاقة بين متغيرين او اكثر، يفترض ان توضع البيانات بنوع معروف من الدوال ومن ثم يتم تحديد افضل دالة للبيانات المعطاة.
- تحليل السلاسل الزمنية Time Series Analysis: إن مشاهدة البيانات عبر الزمن تنتج تحليلاً مفيداً لانه تتم مشاهدة سلوك البيانات عبر الزمن بشكل اساسي . وهذا يعني ان قيم الصفة المميزة للبيانات التي يتم فحصها تكون متغيرة عبر الزمن.
- التنبؤ Prediction : يعد من الادوات التي تجذب الانتباه لانها تتمكن من اعطاء مغزي التوقع الناجح في سياق العمل لذا فانه يمكن النظر الي العديد من تطبيقات تنقيب بيانات العالم الحقيقي كانها تنبؤ بحالة بيانات مستقبلية معتمدة علي بيانات سابقة وحالية.

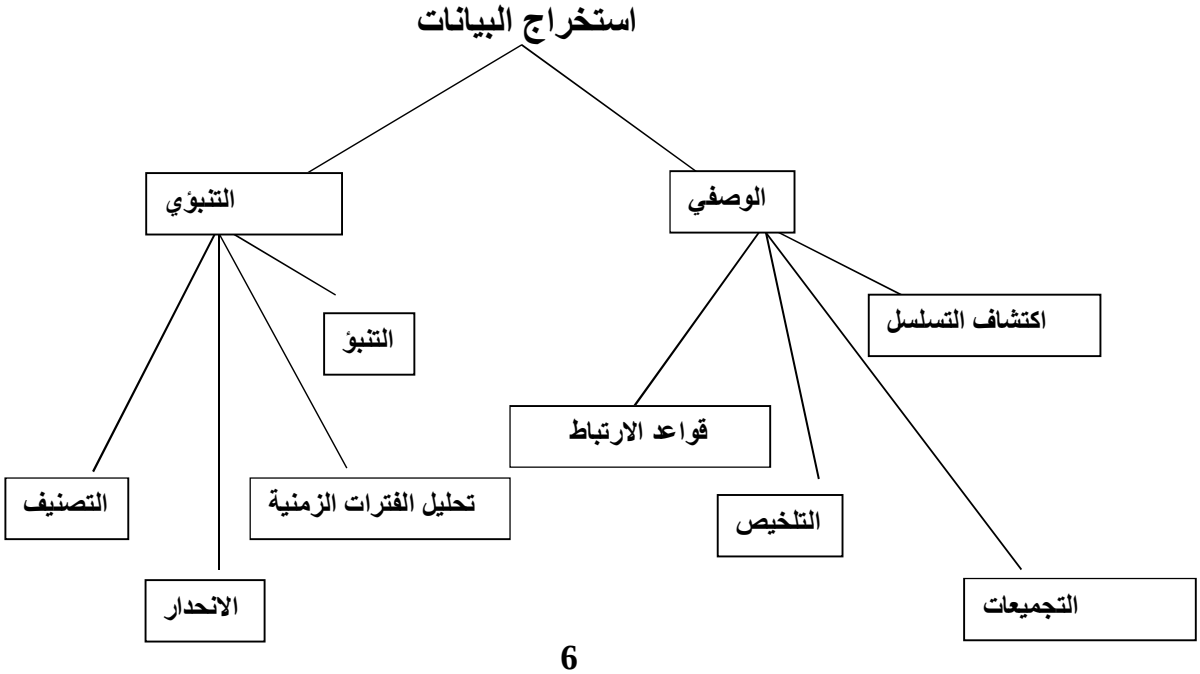
ثانياً: النموذج الوصفي Descriptive Model : وهو النموذج الذي يعرف الانماط والعلاقات في البيانات ، وعلي العكس من النموذج التنبؤي، فان النموذج الوصفي يستخدم كطريقة لاستكشاف خصائص البيانات التي تتم دراستها وليس للتنبؤ بخصائص جديدة ، ويتضمن هذا النموذج اشهر ادوات تنقيب البيانات.

- العنقدة Clustering : وهي عملية تنظيم البيانات في اصناف.
- التلخيص Summarization: ويسمي أيضاً بالمميزات (او الخواص) Characterization او العمومية Generalization، ويمكن تعريفها علي انها وصف الخصائص العامة للنماذج.
- قواعد الربط Association Rules : إن تحليل الربط يشير بصورة بديلة الي تحليل العلاقة لاكتشاف ما هو الشئ المشترك بين البيانات.
- المرئية Visualization: إن مرئية النتائج يكون مساعداً في تسهيل ملاحظات مخرجات خوارزميات تنقيب البيانات وفهمها.
- اكشتاف التسلسل Sequence Discovery : إن تحليل او اكتشاف التسلسل يستخدم لتحديد أنماط متسلسلة في البيانات وهذه الانماط معتمدة علي تسلسل زمن التأثيرات. [4] الشكل (1) يوضح نماذج تنقيب البيانات.

2.6 أهداف التنقيب في البيانات(Data mining)

2.6.1 هناك ثلاثة أهداف للتنقيب في البيانات:(Data mining)

- من أجل تعليل بعض الظواهر المرئية. مثال: لماذا زادت نسبة المدخنين في الوطن العربي؟
- من أجل التثبت من نظرية ما. مثال: التثبت من النظرية التي تقول بأن الأسر الكبيرة تهتم بالضمان الصحي أكثر من الأسر الصغيرة عددا.
- من أجل تحليل البيانات للحصول على علاقات جديدة وغير متوقعة, مثال: كيف سيكون الانفاق العام إن كان ملازما لعمليات خداع واسعة من قبل البطاقات الائتمانية. [10]



الشكل 2.1: استخراج البيانات النموذج و المهام

2.7 العنقدة

هي عملية استخلاص او ايجاد مجموعة من الاشياء (الكائنات) المتشابهة فيما بينها بشكل مجاميع ، بحيث تكون عنصر كل مجموعة متشابهة مع بعضها بصورة اكبر وهذه الاشياء في العنقود الواحد تكون متشابهة فيما بينها ومختلفة عن العنقود الاخر. وان قوة تشابه الاشياء في العنقود الواحد تؤدي الي افضل عنقدة ودالة التشابه تعد مقياساً لجودة العنقدة.
تعتمد معظم خوارزميات العنقدة علي افتراضات مستمدة من تعريف المجاميع الجزئية والمتمثلة او الموجودة في مجموعة البيانات تحت الاختبار واغلب الخوارزميات تعمل بدون مشرف.
يعد تحليل العنقود من الاهداف المفيدة في تحديد التوزيعات البيانية الرغوب فيها وتحديد انماط للبيانات الاساسية. [6]

2.8 خوارزمية العنقدة K-Means Clustering:

تستخدم هذه الخوارزمية لعنقدة او تجميع بعض الكائنات اعتماداً علي خصائصها الي k عنقود، وتتم عملية العنقدة من خلال تقليل المسافات بين الكائنات ومركز العنقود.
وهي واحدة من ابسط الخوارزميات التي قدمت الحل لمعضلة تجميعات معروفه. الاجراء يتبع طريقة سهلة وبسيطة لتصنيف مجموعة البيانات وذلك من خلال عدد محدد من التجميعات(افترض س تجميعة) محدده مسبقا. الفكرة الاساسية تحديد ص كمتوسط، واحد لكل عنقود. هذه المتوسطات يجب ان تحدد بطريقة مناسبة لان وضع هذه الاوساط في اماكن مختلفة يسبب نتائج مختلفة، لذلك انسب خيار هو ان تضعهم ابعد ما يمكن من بعضهم. الخطوة التالية تاخذ كل نقطة تنتمي الي مجموعة البيانات المعطاة و ربطها باقرب نقطة وسطية. عندما تربط كل النقاط، المرحلة الأولي اكتملت. في هذه المرحلة نحتاج لنعيد حساب المتوسط ص مرة اخري من منتصف التجميعات الناتجة من المرحلة السابقة. بعد الحصول علي هذه المتوسطات الجديدة، يتعين القيام بالربط بين نفس نقاط مجموعة البيانات و اقرب متوسط جديد. تم انشاء الحلقة. ونتيجة لهذه الحلقة نجد ان المتوسطات ص تغير أماكنها في كل مرحلة حتي لا يكون هناك تغيير. بمعني أخر حتي تتوقف المتوسطات عن تغيير اماكنها.[7]

خطوات الخوارزمية

مدخلاتها وهي

- $D = \{d1, d2, \ldots, dn\}$ حيث N عدد نقاط البيانات.
- K عدد العناقيد المرغوب.

ومخرجاتها هي

- مجموعة من العناقيد عددها K .

خطواتها وهي

1. تحديد k مركزاً Centroids للعناقيد بشكل عشوائي.
2. حساب المسافة بين كل نقطة وبين جميع المراكز، ويتم ذلك باستخدام البعد الاقليدي. يعطي البعد الاقليدي d_{ij} حسب العلاقة التالية:

$$d_{ij} = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2}$$

حيث : n أحداثيات نقاط البيانات.
Xik إحداثيات الخاصية K للنقطة i.
xjk إحداثيات الخاصية K للنقطة j (يكون في العادة إحداثيات المركز).

3. تجميع الكائنات لاقرب مركز لها.
4. حساب المتوسط الجديد لكل عنقود.
5. تكرار الخطوات من 2 الي 4 حتي حصول الاستقرار (عدم وجود كائنات تنتقل ضمن التجمعات) أو حتي التكرار لعدد محدد من المرات.
6. نهاية الخوارزمية.

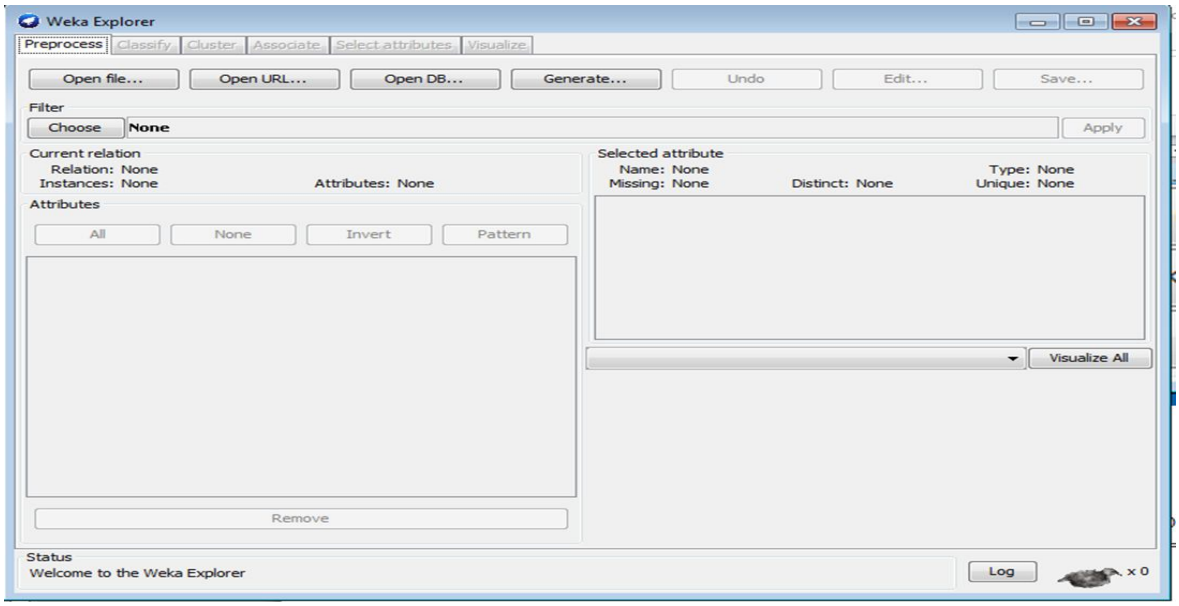
يعتمد أداء هذه الخوارزمية علي المواقع الاولية لمراكز التجمعات، ومن المفضل تنفيذ هذه الخوارزمية عدة مرات مع اختلاف المراكز في كل مرة. وتندرج هذه الخوارزمية تحت الهدف بدون اشراف لانها تصنف النقطة تلقائياً اعتمادا علي شرط معين (مثل اقرب مسافة الي مراكز التجمع) [9].

2.9 تحديد حزمة البرمجيات

اختيار انسب حزمة برمجيات ليست بالامر السهل نظرا بالعودة الي الابحاث السابقة،لان حزم البرمجيات لديها نقاط قوة و ضعف. اعتمادا عليالمعلومات المعروفة لدراسة وتصنيف فئات الفقر بالسودان تم إختيار، Weka 3.6.9 و التي تم تضمينها في الجافا مع اخر نسخة من نظام التشغيل windows 7 مع المعالج Intel Core 2Quad@2.83 GHz و الذاكرة بحجم 2 قيقا ، مع استخدام معالج ماكرو بسيط الي حد كبير كبرنامج يسمح باستيراد البيانات مباشرة من اكسل. لان كل البيانات التي يتم تجميعها من اي مصادر يمكن تنزيلها في اكسل ،لانه يعد من البرامج الشائعة والمستخدمة علي نطاق واسع.

WEKA 2.10

Waikato Environment for knowledge Analysis بيئة وايكاتو لتحليل المعرفة وهي احد فروع الذكاء الاصطناعي حزمة مشهرة من تطبيقات التعليم الالى مكتوبة بلغة جافا , تم تطويرها في جامعة وايكاتو في نيوزيلندا . Weka من البرمجيات المفتوحة المصدر بموجب رخصة جنو العمومية العامة[8] .. انظر الشكل (2.1).



*الشكل (2.2) الصفحة الرئيسية لتطبيق Weka النسخة 3.6.9

الشكل (2.1) يظهر على الشاشة الافتتاحية مع الخيارات المتاحة. في البداية لا يوجد سوى خيار لتحديد علامة التبويب المعالجة الاولية في أعلى الزاوية اليسرى. ويرجع ذلك إلى ضرورة تقديم مجموعة البيانات إلى التطبيق بحيث المعالجة. بعد أن تم المعالجة الاولية للبيانات هنالك علامات التبويب الأخرى تنشط للاستخدام.

هناك ست علامات تبويب:

2.10.1 وصف التطبيق

يقسم تطبيق Weka من مجموعة الادوات المرتبة واللوغريثمات لتحليل البيانات, ونماذج التنبؤ معا في واجهة مستخدم رسومية تسهيلاً لاستخدامها.

النسخة الاصلية لتطبيق Weka الغير مبنية بلغة جافا تعريق ب TCL/TL وكانت واجهات في الاغلب من طرف ثالث حيث نموذج الخوارزميات بتنفيذ بلغات برمجه أخرى بينما أدوات تجهيز البيانات تستخدم لغة برمجة (c) وضع ملف تنفيذي ليعمل نظام التعليم الالى وصممت النسخة الاصلية أساسا كاداه لتحليل بيانات الجاذبية في البحوث الزراعية . ولكن نسخة تطبيق Weka 3.6.4 الاكثر حداثه والمعتمدة كليا علي جافا . بدأ تطويرها في عام 1997م ,تستخدم الان في الكثير من مجالات التطبيق المختلفة بالاخص للاغراض البحثية والتعليمية.

وهنا بعض المزايا لتطبيق Weka

- متاح مجاناً تحت رخصة جنو العمومية العامة.
- قابلية العمل . منذ أعتمده كليا علي لغة برمجة جافا أصبح يمكن تشغيله علي أغلب المنصات الحاسوبية الحديثة.
- غنى بمجموعة من ادوات تجهيز البيانات وتقنيات النمذجه.

- سهولة الاستخدام لاستخدامه واجهة مستخدم رسومية.

يدعم Weka العديد من المهام القياسية لتنقيب البيانات ويشكل اكثرا تحديدا تجهيز البيانات , المجاميع , التصنيف , التراجع , التصور و أختيار ميزة وتعتمد جميع تقنيات Weka علي فرضية ان البيانات المتاحة هي ملف أو علاقة واحدة حيث يتم وصف كل نقطة من البيانات بعدد محدد من السمات (عادية) سمات رقيمة او رمزية , ويدعم أيضاً بعض انواع السمات الاخرى.

ويوفر تطبيق Weka اتصال بقواعد البيانات SCL بأستخدام أتصال قواعد بيانات جافا ويمكنه معالجة النتيجة التى تم ارجاعها عن طريق استعلام قاعدة البيانات . وهي ليست مقتدرة لتنيقب البيانات متعدد العلاقات.

ولكن هناك تطبيق منفصل لتحويل مجموعة مرتبطة من جداول قاعدة البيانات لجدول واحد ولذا يكون مناسباً لمعالجته ب Weka وهناك مجالا مهما لم يغطى حالياً بالخوارزميات المضمنة في توزيعه تطبيق Weka وهي النمذجة التتابعية.

واجهة المستخدم الرئسية لتطبيق ال Weka هي المتصفح .ولكن في نفس الوظائف الاساسية يمكن الوصول اليه من خلال واجهة مكون تدفق المعرفة وسطر الاوامر.

وهنالك ايضأ التجارب الذي يسمح بالمقارنة المنهجية الأداء التنبؤي لخوارزميات التعليم الالى ب Weka المجمعة في قاعدة بيانات.

ميزة واجهة المتصفح لها عدة واجهات. تتيح الوصول للمكونات الرئيسية في ساحة العمل.

- المعاجة الاولية preprocess تعمل علي استرداد البيانات من قاعدة البيانات ملف بامتداد csv وتستخدم ما يسمى بخوارزمية التصفية لاجراء المعالجة الاولية لهذه البيانات. ويستخدم هذا الفلتر لتحويل البيانات علي سيبل البيانات (تحويل السمات الرقيمة الي منفصلة) وجعلها من الممكن حذف الحالات او السمات وفقاً لمعايير محددة.
- التصنيف classify: تمكن لوحة التصنيف المستخدم لاستخدام خوارزميات التصنيف والتراجع(وتسمى عرضا بالمصنفات في تطبيق ال WEKA) لنتائج مجموعة من البيانات لتقدير دقة نتائج نموذج التنبؤ , وتصور التنبؤات الخاطئة ومنحنيات ROC وما الي ذلك . او النموذج نفسه (اذا كان النموذج قابل للتصور علي سبيل المثال شجرة القرار).
- التجميع CLUSTER: تعطى لوحة التجميع حق الوصول لتقنيات التجميع في تطبيق WEKA علي سبيل المثال خوارزمية Simple K-means algorithm. وهناك ايضاً تنفيذ خوارزمية تحقيق اقصى التوقعات لتعليم مزيج من التوزيعات الطبيعية.

- المعاونAssociation: توفر لوحة المعاون حق الوصول للاجراءات المجتمعة للمتعلمين التى تحاول ان تعرف كل العلاقات المهمة بين السمات في البيانات .
- تحديد السماتAttribute selection: توفر لوحة تحديد السمات خوارزميات لتحديد أغلب السمات المتوقعة في مجموعة البيانات.
- التصورVisualize: تعرض لنا لوحة التصور مصفوفة من الاحداث المبعثرة حيث ان الاحداث المبعثرة يمكن اختيارها وتكبيرها وتحليلها باستخدام مزيد من المشغلين المختارين.

2.10.2 العنقدة في تطبيق Weka :

يحتوى تطبيق Weka علي التجميع للعثور علي الحالات المتشابهة في مجموعة البيانات. ومجموعة المخططات المتاحة في تطبيق ال Weka هي Cobweb,EM,K-Means , and FarthestFirst,X-means, . ويمكن تصور التجميع ومقارنتة الي تجميع حقيقي (لو اعطى) ويستند التقييم علي سجل احتمال اذا انتج مخطط التجميع وتوزيع الاحتمالات[6].

2.10.3 تنقية البيانات:

قبل تجميع البيانات، يتم تنفيذ مهام التنقية التالية على البيانات:

1- ملء البيانات المفقودة.
2- تحويل البيانات الي الصيغة القياسية(csv).
3- إصلاح الاخطاء والانحرافات,القيم المتطرفة.
4- ازالة البيانات المشوهة.

منهجية البحث والبيانات

3.1 مقدمة

تمت عمليات الحصر الشامل للفقراء للعام 2011م ابتداء من مرحلة التحضير التي شملت اجتماعات لاعداد خطة شاملة وممرحلة ومفصلة لكل الانشطة لاجراء الحصر للوصول لكل الاسرة الفقيرة.

وفر الحصر بيانات وإحصاءات ومؤشرات عن الاسر الفقيرة. وهي احصاءات لاغني عنها في مجال التخطيط وتقيم متابعة التصدي للفقر,هذه البيانات تم تبويبها علي حسب القطاعات مما ساعد في تحديد المؤشرات.

وشملت البيانات المستخلصة من الحصر علي معلومات الحالة التعليمية للفقراء كالامية والمحصلة التعليمية والالتحاق براحل التعليم المختلفة.

كما وفر الحصر بيانات عن التركيبة العمرية والنوعية للسكان ومعلومات عن السكن والمساكن وانواع المباني والخدمات المتوافرة للفقراء من حيث مياه الشرب ووقود الطهي وانواع الصرف الصحي.

وقد توافر من الحصر بيانات عن امتلاك الاسرة الفقيرة للاراضي الزراعية وغير الزراعية ومصادر معيشة الاسرة وبيانات عن الحالة العملية والمهن والقطاعات الاقتصادية التي تعمل فيها الاسر الفقيرة وهذه المعلومات مهمة جداً لاغراض تحليل وتحديد عناصر الحياة الطيبة علي مستوى الاسرة.

3.2 اهداف الحصر:

ان الهدف الاستراتيجي للحصر الشامل للفقراء وهو ايجاد قاعدة بيانات واسعة تمكن من وضع استراتيجية وضحة المعالم للحد من الفقر ومناهضته ولذلك جاءت عمليات الحصر الشامل للفقراء لتعين واضعي السياسسات والباحثين في:

- تحديد وتصنيف الاسر الفقيرة في الاحياء والقري والفرقان الي فئات متدرجة..
- بناء قواعد معلومات للاسرة الفقيرة ..
- توفير بيانات عن الاسر الفقيرة تعين والباحثين وواضعي السياسات في دراسة اليات توليد الفقر واستدامته.
- تحديد خرائط الفقر
- مراحل تنفيذ حصر الاسر الفقيرة:
- يحتاح الي حصر الي لجان متخصصة للقيام بالعمليات الميدانية والفنية والمالية.تكونت لجنة عليا للحصر برئاسة السيد/ نائب الامين العام لديوان الزكاة وعضوية عدد من المختصين والخبراء اشرفت علي كافة مراحل الحصر وقد تحددت مهامها بالاتي:
- تحديد من هو الفقير
- تحديد الكيفية التي يعرف بها الفقير

- تحديد المرتكزات الاساسية التي سيبني عليها الحصر الشامل وفقاً لمرجعية الدراسات الاحصائية والقياسية للفقر في السودان.
- تحديد الاجراءات الفنية اللازمة للحصر الميداني
- اجازة العمليات الفنية المطلوبة لاستخلاص النتائج.
- وضع اوزان لتصنيف الفقراء لمجوعات

3.3 تصميم الاستمارة:

اعدت استمارة حصر الاسر الفقيرة بواسطة اكاديميين وخبراء احصائيين واحتوت الاستمارة علي 51سؤالاً في خمسة اقسام وهي:

- خصائص افراد الاسرة
- دخل الاسرة من النشاط الزراعي (بشقية النباتي والحيواني) ودخل الاسرة من الانشطة غير الزراعية.
- المستوي المعيشي للاسرة
- خبرات المبحوت مع مؤسسات الرعاية الاجتماعية
- خصائص المسكن

ملحوظة: تم ارسال هذه الاستمارات الي الولايات لتحديد الاسره الفقيرة وتم والتدريب عليها..

3.4 الحصر الفعلي:

تم جمع البيانات والتي تمت بطريقة المقابلة الشخصية بواسطة باحثين للاسرة الواحدة وذلك لضمان عدم التحيز. ومن مزايا المقابلة الشخصية انها طريقة تعطي الباحث وسيلة اخري لجمع المعلومات وهي المشاهدة او الملاحظة بالعين المجردة. بعد ان تم جمع الاستمارات تمت مراجعتها للتاكد من الاخطاء التي حدثت اثناء جمع البيانات مثل(التاكد من خلوها من التناقض)

3.5 المشاكل التي واجهت عمليات الحصر:

- وسائل الحركة
- مشاكل فنية متمثلة في عدم وضوح ارقام الاستمارات.
- مشاكل تنظمية : معروف ان اي حصر يتم بواسطة لجان الزكاة القاعدية وجد ان هناك ضعف وعدم قدرة لجان الزكاة علي تنظيم اعملها اعاق القيام بعمليات الحصر بالمستوي المطلوب.
- ضعف الانشار الجغرافي ادي الي هدر وقت طويل في عمليات المراجعة
- قصر فترة التدريب ادي الي حدوث بعض الخلل مثل(عدم تعبئة بعض اسئلة بالطريقة الصحيحة, وجود عدد من الاسئلة لم تتم الاجابة عنها)

3.6 العمليات الفنية ومعالجة البيانات:

تحليل بيانات الحصر الشامل للفقراء يعد الخطوة الاولي نحو صياغة استراتيجية لتخفيف ظاهرة الفقر. فالدخل المتدني لا يعتبر السمة الوحيدة للفقر فعادة ما يرتبط الفقر بسوء التغذية ومستويات التعليم المتدنية والظروف المعيشية المتواضعة وصعوبة الحصول علي الخدمات الاساسية مثل

التعليم والصحة والمياه ومن ثم عملية توزيع الخدمات الاجتماعية في السودان لابد ان نرتكز علي الاعداد الفعلية للسكان الفقراء بالاضافة الي خصائص هؤلا السكان الذين يعيشون تحت خط الفقر. وتزداد اهمية هذا النوع من التحليل بالنسبة لصانعي السياسيات الذين تقع علي عاتقهم اعداد واستهداف استراتيجيات للقضاء علي الفقر.

استند الحصر علي التعاريف المعتمدة وطنياً بشان الاسرة(التعريف الاحصائئ) كما اعتمد تعريف الاسر الفقيرة اذا انطبقت عليها واحدة او اكثر من الصفات التالية:

- الا يكون للاسرة دخل ورب الاسرة عاطل عن العمل وليس فيها من له القدرة علي العمل
- ان يكون دخل معيل الاسرة اقل من 120 في الشهر وليس للاسرة دخل من اي مصدر اخر
- ان يكون متوسط استهلاك الفرد في الاسرة اقل من 114 جنيه في الشهر
- ان يكون رب الاسرة عاطلا عن العمل جبرا نتيجة لعجز او مرض او عدم وجود عمل.
- ان يكون مجموع دخل الاسرة في اقل من الحد الادني للاجور.
- الاسرة الموبوءة بالامراض المكلفة مالياً ورب الاسرة يعمل باجر
- رب الاسرة المعاشي اذا كان لديه مرض مزمن وله اسرة مكونة من ستة افراد وجميعهم في مراحل التعليم وليس لديه اي مصدر دخل اخر.
- رب الاسرة الذي يملك اصولا مثل منزل او ارض زراعية او عربة اجرة لا تعمل. وكانت هذه الاصول لا تدر دخلاً ولا يملك مالا لاستثمارها ولديه اسرة تعتمد عليه في معاشها.
- اصحاب الاعمال الحرة كالنجارين والمزارعين والحدادين الذين لا ينتجون كفايتهم من القوت ولا دخل اخر لهم.
- العمال الزراعيون والرعاة الذين لا يملكون مواشي و لا دخل اخر لديهم، ولديهم اسر تتكون من ستة افراد فاكثر.

3.7 مرحلة بناء الاوزارن وحساب الدرجات للاسرة الفقيرة

معلوم ان الفقراء لا يتساون في فقرهم فهناك من هم دون حد الكفاف وهناك من يبلغون نصف الكفاية وهناك من يبلغون اكثر من نصف الكفاية وطبيعي بعد ان تم حصر الفقراء بالعمل الميداني تم ترتيب الفقراء حسب درجات فقرهم باستخدام الاوزان.

3.8 كيفية إعداد الاوزان:

بناء علي التعريف الاجرائي للفقير والمسكين واعتبار ان الفقر ظاهرة متعددة الابعاد رأت اللجنة الفنية في معهد علوم الزكاة استخدام اثني عشر مؤشرا وفرها الحصر الشامل لديوان الزكاة يقيس كل واحد منها بعداً من ابعاد الفقر، مؤشرين لقياس الانفاق والتكيف مع نقص الدخل ومؤشر واحد يقس بعد الصحة لدلالة وجود شخص او اكثر مريص مرضاً مزمنا بين افراد الاسرة ومؤشر يقيس

التعليم وبقية المؤشرات تقيس مستويات المعيشة مثل ملكية المسكن ونوع مادة بناء المسكن ومصدر الاضاءة ونوع وقود الطهي الغالب ومصدر مياه الشرب وأعطيت أوزان لكل مؤشر تتراوح بين خمس الي عشر درجات تم توزيع الدرجات علي الاسر من الافضل الي الاقل ثم الاقل وبالتالي تم تصنيف الفقراء حسب الدرجات الي ثلاث فئات كما يلي:

جدول رقم (3.1) يوضح تصنيف فئات الفقر

الفئة	الدرجات التي تحصلت علها الاسرة
لفئة الأولي (الأشد فقراً)	33-1
الفئة الثانية	66-34
الفئة الثالثة	67 فأكثر

❖ **درجات المبحوثين من الفقراء والمساكين:**

- **عدد افراد الاسرة منسوبا الي عدد الغرف:**

جدول رقم (3.2) يوضح درجات عدد افراد الاسرة منسوبا الي عدد الغرف

عدد الافراد في الغرفة	الدرجة
لا توجد غرف	1
7 فاكثر في الغرفة	2
6-4 في الغرفة	3
3-2 في الغرفة	4
فرد في الغرفة	5

التبرير: تعتبر الدار الواسعة في مساحتها وعدد غرفها من مظاهر الغني اد ترتبط طرديا مع حالات الدخل وظروف الاسرة.

- **مادة بناء المنزل:**

جدول رقم (3.3) يوضح مادة بناء المنزل

الحالة	الدرجة
اسمنت	10
طوب واسمنت	9
طوب محروق	7
جالوص وطوب	4
جالوص	3
قش وعيدان	2
اخري	1

التبرير: يشار هنا الي انه علي الرغم من البيئات المختلفة قد تستوجب بناء من مواد معينة كالبيئة المطرية و نوع التربة تجعل المنازل مبنية من القش و العيدان.وعليه يظل الحكم العام بان المواد المكونة لمبني المنزل تشير الي مستوي دخل الاسرة .

- **ملكية المسكن:**

جدول رقم (3.4) يوضح ملكية المسكن

الحالة	الدرجة
ملك	10
حيازة	8
تركة	5
مع اسرتي	4
مع اسرة شريكي الزوجي	3
ايجار	2
اخري	1

التبرير: لان ملكية المسكن تؤثر علي الدخل الحقيقي للاسرة بطريقة مباشرة ام غير مباشرة فمن له ملك او حيازة فهو يوفر تكلفة الايجار وبالتالي يزيد من دخله الشهري بمقدار اجر المثل وعليه فانه بحسب مستويات الوفرة او الاقتطاع من دخل الاسرة المخصص لاستهلاكها وضعت الاوزان علي الترتيب الوارد اعلاه.

- **الصرف الصحي:**

جدول رقم (3.5) يوضح الصرف الصحي

الحالة	الدرجة
سايفون	10
مصاص	7
حفرة	5
جردل	3
اخري	2
لايوجد	1

التبرير: من الواضح ان هذا الترتيب يبدو مناسباً لحالات (نوع) المرحاض وبني اساساً علي امرين:

- الاول: قدرة الاسرة ومواردها.
- الثاني: زيادة الوعي فيها وهو نتاج للعديد من العوامل مثلا التعليم والبئية ومستوي الدخل في المنطقة ..الخ.

- **مصدر مياه الشرب:**

جدول رقم (3.6) يوضح مياه الشرب

الحالة	الدرجة
ماسورة بالمنزل	10
ماسورة جيران	6
ماسورة عامة	5
كاروه مياه	2
اخرى	1

التبرير: حيث اخذ في الاعتبار احتياجات الاسرة من المياه الامنة. وقدرة الاسرة علي توفيرها بحسب البئية ومستوي الوعي في الاسرة والمنطقة علي التوالي وهذا من ناحية ومن ناحية اخرى فان استهلاك المياه الامنة يعتبر احد مؤشرات التنمية.

- **نوع الإضاءة:**

جدول رقم (3.7) يوضح نوع الاضاءة

الحالة	الدرجة
كهرباء عامة	10
مولد بالمنزل	8
مولد بالحي	6
لمبة كيروسين	4
شمع	3
اخري	2
لايوجد	1

- **وقود الطهي:**

جدول رقم (3.8) يوضح وقود الطهي

الحالة	الدرجة
كهرباء	10
غاز	8
كيروسين	6
فحم	4
حطب	3
اخرى	1

- **التكيف مع الفقر:**

جدول رقم (3.9) يوضح التكيف مع الفقر

الحالة	الدرجة

لا ينطبق	10
اخري	9
ابحث علي اي عمل	8
استدين	5
الجأ الي انواع اقل تكلفة	4
ابيع بعض الاشياء	3
الجأ الي اخرين للحصول علي موارد الاكل والشرب	2
اتنازل عن بعض الوجبات	1

التبرير: كيفية الوفاء بالاحتياجات عنصر اساسي في بيان القدرة والمرتبطة باهم انواع الفقر وهو الفقر البشري. فكلما كانت الحلول التي يلجأ اليها الشخص ضعيفة او غير مناسبة مع بيئتة كلما دل ذلك علي عدم القدرة او الاستطاعة والتي هي عنوان اساسي للفقر البشري.

- **الصحة:**

جدول رقم (3.10) يوضح بعد الصحة

الحالة	الدرجة
اذا كان لا يوجد مرضي في الاسرة	10
اذا كان عدد المرضي في الاسرة(1-2)	5
اذا كان عدد المرضي في الاسرة(3-4)	3
اذا كان عدد المرضي في الاسرةر5 فما فوق	1

التبرير: تعتبر الصحة مؤشر اساسياً للتنمية المحلية والاجتماعية ويرتبط بها قدرة الانسان علي المعالجة وتعاطي الدواء ولذلك تعتبر عوامل الافقار في الوقت الراهن بحيث كلما زاد عدد المرضي في الاسرة قلت قدرتهم علي تخصيص مبلغ اكبر من ميزانية الاسرة لاستهلاك السلع والخدمات الاخرى.

- **عدد افراد الاسرة:**

جدول رقم (3.11) يوضح عدد افراد الاسرة

الحالة	الدرجة
اذا كان عدد افراد الاسرة(1-3)	5
اذا كان عدد افراد الاسرة(4-6)	3
اذا كان عدد افراد الاسرة 7 فما فوق	1

التبرير : وهو مؤشر يرتبط عكسياً مع المؤشر الخاص بنسب افراد الاسرة في كل غرفة بحيث يعمل هذان المؤشران لايجاد دقيق للوزن الخاص بالاسرة في هذا الجانب.

- **التعليم لرب الاسرة:**

جدول رقم (3.12) يوضح التعليم لرب الاسرة

الحالة	الدرجة
جامعي	5

ثانوي	4
ابتدائي	3
يقرأ ويكتب	2
أمي	1

- **قيمة الاستهلاك:**

جدول رقم (3.13) يوضح قيمة الاستهلاك

الحالة	الدرجة
3-1	1
6-4	2
9-7	3
13-10	4

3.9 تحليل البيانات:

للتحقق من تغطية اهداف الحصر الشامل للفقراء تم استخدام بعض الاساليب الاحصائية الوصفية مثل التكرارات والنسب المئوية والمتوسطات ثم استخدام الجداول التقاطعية للمقارنة بين متغيرات الخصائص الديموغرافيه للاسرة الفقيرة والافراد, التعليم والصحة والخصائص الاقتصادية وخصائص المسكن والانفاق الاستهلاكي للاسر الفقيرة والتكيف مع الفقر.

تم استخدام برنامج التحليل الاحصائي في تحليل بيانات الحصر الشامل [3].

بعد ان اكتمل حصر الاسر الفقيرة وتم تقسيمهم الي ثلاثة فئات وتحديد الاسر الفقيرة باستخدام الاوزان, استخدمنا هذه البيانات في برنامج Weka باستخدام خوارزمية العنقدة K-means لتصنيف فئات الفقر بعمل اكثر من تجربة والهدف من ذلك هل التصنيف المتبع في ديوان الزكاة مثالي ام هناك تصنيف اخر.

المناقشة والنتائج

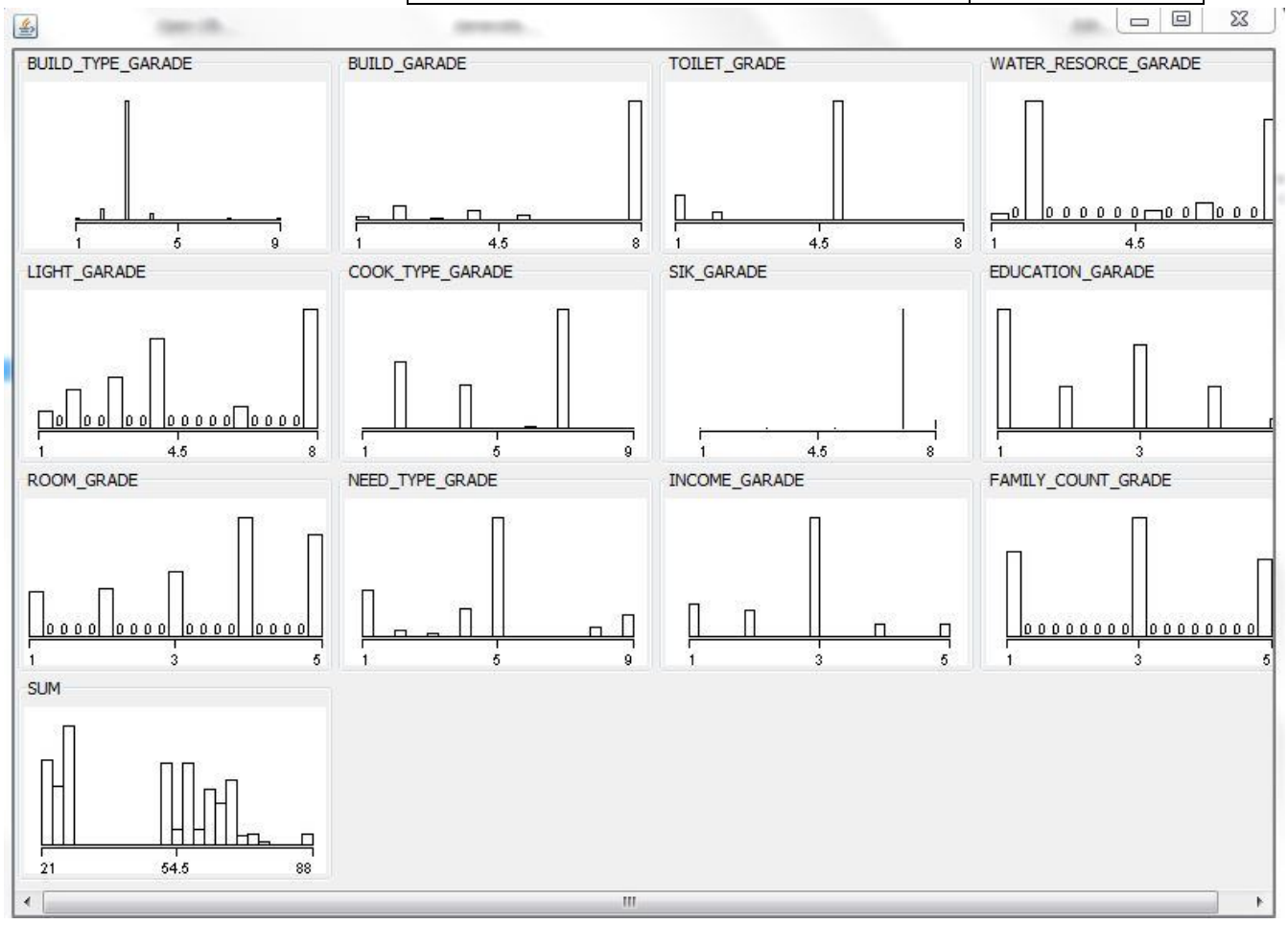
4.1 وصف البيانات

جدول رقم (1)، يلخص البيانات المستخدمة في هذه التجربة. الشكل 3 يظهر خصائص البيانات وتوزيعها.

جدول رقم (4.1) يوضح وصف البيانات المستخدمة

عددها	14868	**خصائص البيانات**	رقمية
عدد الخصائص	13	**خصائص السمات**	حقيقي
القيم المفقودة	لايوجد	**المهام المتعلقة**	العنقدة

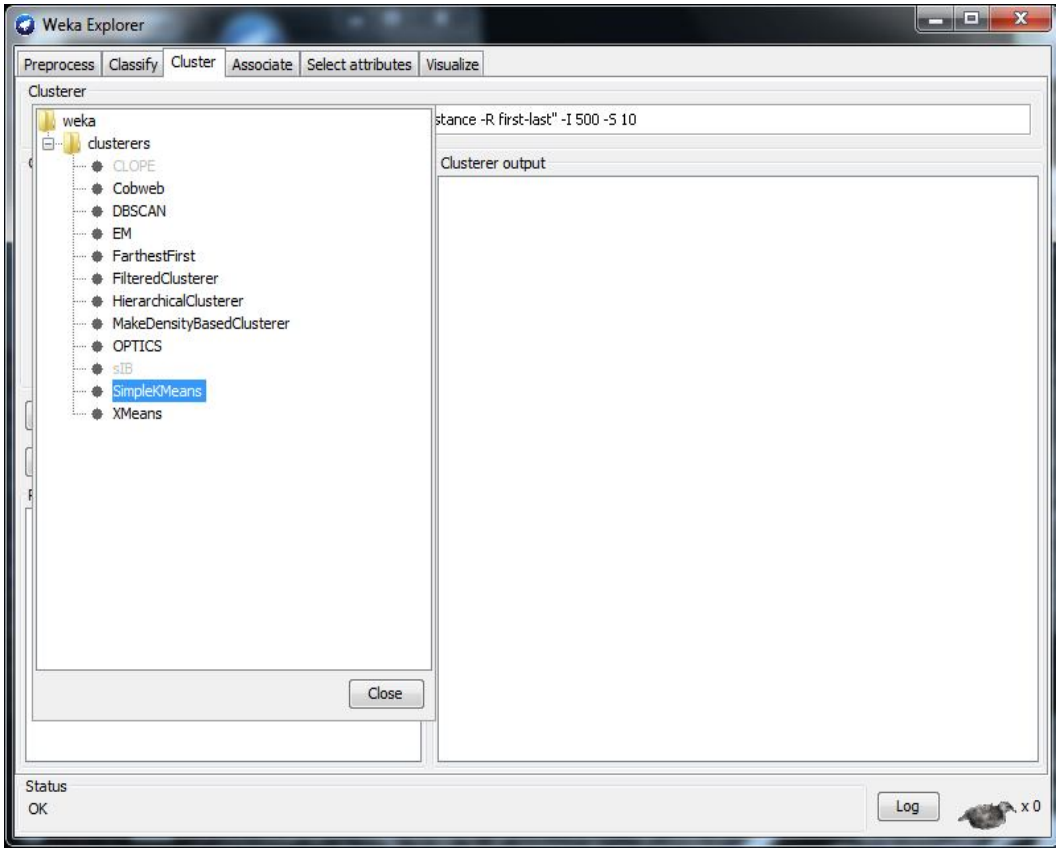
اسماء الصفات			
نوع بناء المنزل	**وقود الطهي**	**نوع الاحتيجات**	
نوع المرحاض	**نوع المرض**	**دخل الاسرة**	
مصدر المياه الشرب	**نوع التعلم**	**عدد الاسرة**	
نوع الاضاءة	**عدد الغرف**	**نوع ملكية المسكن**	
المجموع			



الشكل 4.1: تصور لخصائص البيانات وتوزيعها

4.2: وصف التجارب

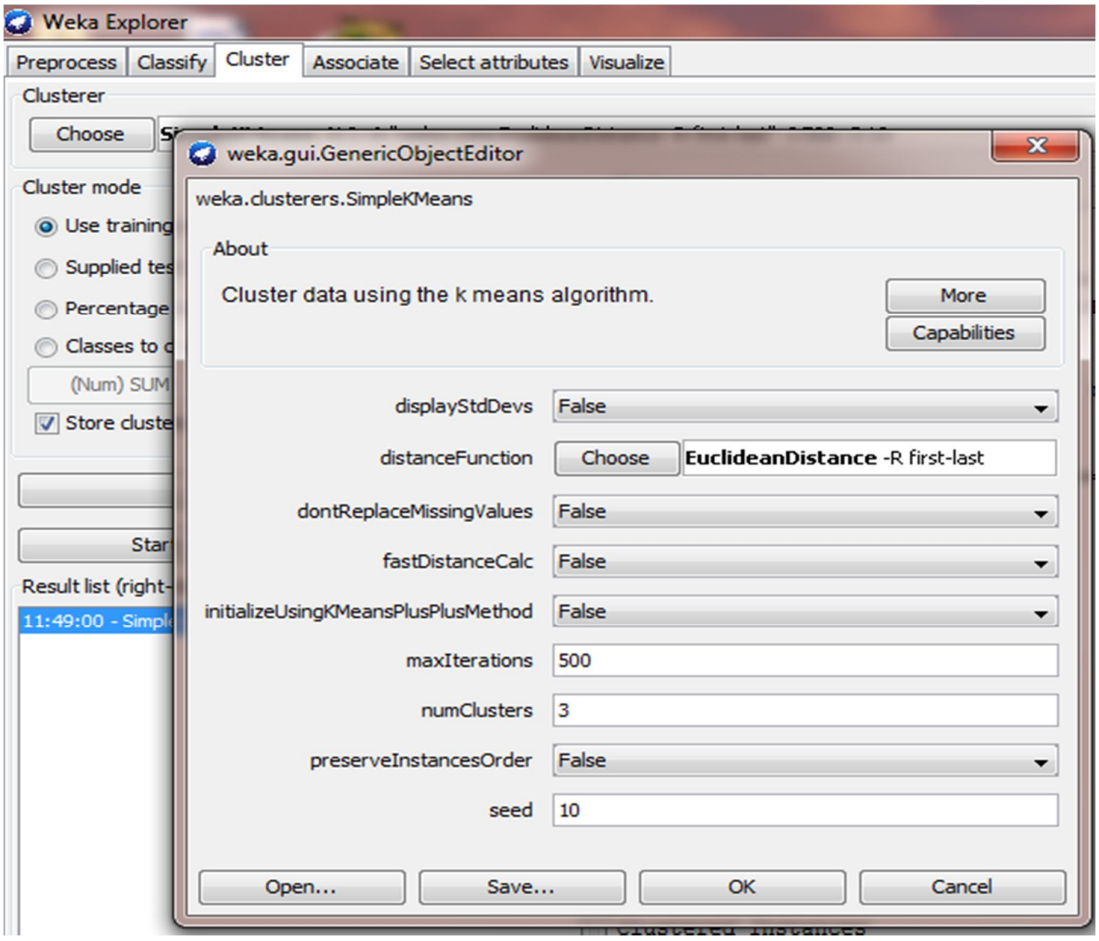
وبالنظر إلى مجموعة البيانات باستخدام ستة تجارب أجريت. ونظرا لوصف هذه التجارب ومناقشاتهم في الأقسام التالية.



الشكل 4.2: صورة لصفحة التجميع في برنامج الويكا

4.3 التجربة الاولي

في هذه التجربة قد استخدمنا Weka لتجميع البيانات باستخدام ثلاث عناقيد. انظر الشكل(4.3) عناقيد التجربة، ويتضمن جدول رقم (4.2)نتيجة هذه التجربة.



الشكل 4.3: عناقيد التجربة

جدول رقم (4.2) يوضح نتيجة التجربة الاولي

الخاصية	كل البيانات 14868	العنقود 0 5856	العنقود 1 4980	العنقود 2 4032
بناء المنزل	3	3.3	3.1	2.9
ملكية المسكن	6.9	7.1	6.9	6.7
الصرف الصحي	3.9	4.8	4.2	3.7
مصدر المياه الشرب	2.7	7.7	4.7	3
نوع الإضاءة	3.5	7.4	5	3.4
وقود الطهي	4.2	6.5	5	3.8
الصحة	7.1	7	7	7.1
نوع التعلم	2.1	2.5	2.3	2
عدد الغرف	2.5	3.6	3.5	4.5
نوع الاحتياجات	4.4	4.5	4.5	4.4
دخل الأسرة	2.7	2.9	2.7	2.6
عدد الأسرة	1.8	3.1	3	4.1
المجموع	46.8	64.8	32.7	38.2

4.3.1.نتائج التجربة:

العنقود (0) يمثل المجموعة الثانية وتحتوي على5856 سجل وهي 39% من نسبة البيانات

العنقود (1) يمثل المجموعة الثانية وتحتوي على4980 **سجل** وهي 33% من نسبة البيانات

العنقود(2) يمثل المجموعة الثانية وتحتوي على4032 **سجل** وهي 27% من نسبة البيانات

4.3.2 ملخص التجربة:

خلاصة التجربة الأولي تظهر لنا

لا توجد عناقيد تمثل المجموعة الثالثة

ظهور عنقودين في المجموعة الثانية وعنقود واحد في المجموعة الاولي

4.4.التجربة الثانية:

في هذه التجربة نزيد عدد العناقيد الي 4

جدول رقم (4.3) يوضح نتيجة التجربة الثانية

الخاصية	**كل البيانات 14868**	**العنقود 0 3010**	**العنقود 1 1819**	**العنقود 2 3718**	**العنقود 3 6321**
بناء المنزل	**3.1**	**3.3**	**2.8**	**3.2**	**3**
ملكية المسكن	**6.9**	**7.3**	**2.3**	**7.1**	**8**
الصرف الصحي	**4.2**	**4.8**	**3.4**	**4.8**	**3.8**
مصدر المياه الشرب	**4.7**	**7.5**	**2.7**	**7.4**	**2.5**
نوع الإضاءة	**5**	**7**	**3.1**	**7**	**3.5**
وقود الطهي	**5**	**6.4**	**3.7**	**6.2**	**3.9**

الصحة	7	7	7	7	7.1
نوع التعلم	2.3	2.5	2.3	2.4	2
عدد الغرف	3.5	2.4	3.5	4.6	3.4
نوع الاحتياجات	4.5	4.6	4.8	4.4	4.3
دخل الأسرة	2.7	2.9	2.8	2.8	2.6
عدد الأسرة	3	1.9	3.1	4.1	2.7
المجموع	46.8	60.7	28.1	65.2	34.7

4.4.1 : نتائج التجربة

العنقود 0 يمثل المجموعة الثانية وتحتوي علي **3010 سجل** وهي 20 % من نسبة البيانات

العنقود 1 يمثل المجوعة الاولي وتحتوي علي **1819 سجل** وهي 12 % من نسبة البيانات

العنقود 2 يمثل المجموعة الثانية وتحتوي علي **3718 سجل** وهي 25 % من نسبة البيانات

العنقود 3 يمثل المجوعة الثانية وتحتوي على **6321** سجل وهي 43 % من نسبة البيانات

4.4.2: ملخص التجربة:

في هذه التجربة لاتوجد عناقيد تمثل المجموعة الثالثة

ظهور ثلاثة عناقيد تمثل المجموعة الثانية وعنقود واحد يمثل المجموعة الاولي

4.5 التجربة الثالثة:

في هذه التجربة زدنا عدد العناقيد الي 5

جدول رقم (4.4) يوضح نتيجة التجربة الثالثة

الخاصية	كل البيانات 14868	العنقود 0 2789	العنقود 1 1675	العنقود 2 3416	العنقود 3 3431	العنقود 4 3557
بناء المنزل	3.1	3.3	2.8	3.2	2.9	3
ملكية المسكن	6.9	7.3	2.2	7	7.9	7.9
الصرف الصحي	4.2	4.8	3.4	4.9	3.4	4.2
مصدر المياه الشرب	4.7	7.8	2.6	7.5	3	2.4
نوع الإضاءة	5	7.1	3.1	7.3	3.2	3.8
وقود الطهي	5	6.5	3.7	6.4	3.4	4.6
الصحة	7	7	7	7	7.1	7.1
نوع التعلم	2.3	2.5	2.4	2.5	1.7	2.3
عدد الغرف	3.5	2.4	3.4	4.6	4.2	2.6
نوع الاحتياجات	4.5	4.6	4.9	4.4	4.3	4.3
دخل الأسرة	2.7	2.9	2.8	2.8	2.4	2.8
عدد الأسرة	3	1.9	3	4.1	4	1.7
المجموع	46.8	61.7	28.1	66.2	36.4	35.3

4.5.1 تحليل النتائج:

العنقود 0 يمثل المجموعة الثانية وتحتوي علي 2789 سجل وهي 19% من نسبة البيانات

العنقود 1 يمثل المجوعة الاولي وتحتوي علي 1675 سجل وهي 11% من نسبة البيانات

العنقود 2 يمثل المجموعة الثانية وتحتوي علي 3416 سجل وهي 23% من نسبة البيانات

العنقود 3 يمثل المجوعة الثانية وتحتوي علي 3431 سجل وهي 23% من نسبة البيانات

العنقود 4 يمثل المجوعة الثانية وتحتوي علي 3557 سجل وهي 24% من نسبة البيانات

4.5.2 ملخص التجربة:

كالتجربتين السابقتين لا توجد عناقيد تمثل المجموعة الثالثة

معظم النتائج تمثل المجموعة الثانية وعنقود واحد يمثل المجموعة الاولي .

4.6 التجربة الرابعة:

في هذه التجربة نقوم بزيادة العناقيد الي 7

جدول رقم (4.5) يوضح نتيجة التجربة الرابعة

الخاصية	كل البيانات 14868	العنقود 0 2772	العنقود 1 921	العنقود 2 3314	العنقود 3 2053	العنقود 4 903	العنقود 5 2399	العنقود 6 2506
بناء المنزل	3.1	3.3	2.7	3.2	3.1	2.9	3	2.9
ملكية المسكن	6.9	7.3	2.4	7	7.9	2.2	8	8
الصرف الصحي	4.2	4.8	3.1	4.8	4.8	3.8	4	3
مصدر المياه الشرب	4.7	7.8	2.7	7.8	2.3	2.7	2.5	3
نوع الإضاءة	5	7.1	2.8	7.3	4.5	3.4	3.4	3.1
وقود الطهي	5	6.5	3.3	6.4	6.1	4.2	3.7	2.8
الصحة	7	7	7	7	7.1	7	7.1	7.1
نوع التعلم	2.3	2.5	2.2	2.4	2.8	2.5	1.9	1.5
عدد الغرف	3.5	2.4	4.3	4.6	3.7	2.5	2.3	4.2
نوع الاحتياجات	4.5	4.6	5.2	4.4	5.1	4.3	3.8	4.2
دخل الأسرة	2.7	2.9	2.7	2.8	3	2.8	2.7	2.3
عدد الأسرة	3	1.9	4.2	4.1	2.8	1.9	1.4	4
المجموع	46.8	61.9	29.3	66.2	53.7	27.2	26.4	31.8

4.6.1 تحليل النتائج:

العنقود 0 يمثل المجموعة الثانية وتحتوي على 2772 سجل وهي 19 % من نسبة البيانات

العنقود 1 يمثل المجوعة الاولي وتحتوي على 921 سجل وهي 6 % من نسبة البيانات

العنقود 2 يمثل المجموعة الثانية وتحتوي على 3314 سجل وهي 22 % من نسبة البيانات
العنقود 3 يمثل المجوعة الثانية وتحتوي على 2053 سجل وهي 14 % من نسبة البيانات
العنقود 4 يمثل المجوعة الاولي وتحتوي على 903 سجل وهي 16 % من نسبة البيانات
العنقود 5 يمثل المجوعة الاولي وتحتوي على2399 سجل وهي 11 % من نسبة البيانات
العنقود 6 يمثل المجوعة الاولي وتحتوي على2506 سجل وهي 17 % من نسبة البيانات

4.6.2 ملخص التجربة:

في هذه التجربة تظهر لنا:
لا توجد عناقيد تمثل المجموعة الثالثة
النتائج تظهر المجموعة الاولي الثانية و ظهور المجموعة الاولي في اربعة عاقيد . هذا يعني ظهور المجموعة الاولي بشكل اكبر من المجموعة الثانية.

4.7 التجربة الخامسة:

في هذه التجربة تقوم بزيادة عدة المجاميع الي 8

جدول رقم (4.6) يوضح نتيجة التجربة الخامسة

الخاصية	كل البيانات 14868	العنقود 0 2496	العنقود 1 920	العنقود 2 2818	العنقود 3 1889	العنقود 4 848	العنقود 5 2160	العنقود 6 2350	العنقود 7 1387
بناء المنزل	3.1	3.3	2.6	3.2	3.1	3	3	2.9	3.1
ملكية المسكن	6.9	7.2	2.3	6.9	7.9	2.2	8	8	7.6
الصرف الصحي	4.2	4.8	2.9	4.9	4.8	3.9	4	3	4.4
مصدر المياه الشرب	4.7	7.8	2.5	7.8	2	2.8	2.4	2.3	7.2
نوع الإضاءة	5	7.6	2.8	7.9	4.6	3.4	3.5	3.1	3
وقود الطهي	5	6.6	3.2	6.5	6.1	4.2	3.8	2.8	4.8
الصحة	7	7	7	7	7.1	7	7.1	7.1	7
نوع التعلم	2.3	2.6	2.3	2.5	2.8	2.4	2	1.6	1.8
عدد الغرف	3.5	2.4	4.2	4.6	3.8	2.5	2.2	4	4
نوع الاحتياجات	4.5	4.6	5.2	4.4	5	4.3	4	4	4.5
دخل الأسرة	2.7	3	2.8	2.8	3	2.8	2.7	2.3	2.6
عدد الأسرة	3	1.9	4.1	4.1	3	1.8	1.2	3.9	3.6
المجموع	46.8	62.6	28.2	67.5	53.8	27.1	27.7	27.7	53.2

4.7.1 تحليل النتائج

العنقود 0 يمثل المجموعة الثانية وتحتوي 2496 سجل وهي17 % من نسبة البيانات

العنقود 1 يمثل المجوعة الاولي وتحتوي على920 سجل وهي 6 % من نسبة البيانات

العنقود 2 يمثل المجموعة الثالثة وتحتوي على2818 سجل وهي 19 % من نسبة البيانات

العنقود 3 يمثل المجوعة الثانية وتحتوي على1889 سجل وهي 13 % من نسبة البيانات

العنقود 4 يمثل المجوعة الاولي وتحتوي على848 سجل وهي 6 % من نسبة البيانات

العنقود 5 يمثل المجوعة الاولي وتحتوي على 2160سجل وهي 15 % من نسبة البيانات

العنقود 6 يمثل المجوعة الاولي وتحتوي على2350 سجل وهي 16 % من نسبة البيانات

العنقود7 يمثل المجوعة الثانية وتحتوي على1387 سجل وهي 9 % من نسبة البيانات

4.7.2 ملخص التجربة:

في هذه التجربة تظهر لنا:

- ظهرت لنا العناقيد الثلاثة
- ظهور المجموعة الثالثة في عنقود واحد
- النتائج تظهر المجموعة الاولي و الثانية و ظهور المجموعة الاولي في اربعة عاقيد . هذا يعني ظهور المجموعة الاولي بشكل اكبر من المجموعة الثانية.

.

4.8 التجربة السادسه:

في هذه التجربة تقوم بزيادة عدة المجاميع الي 10

جدول رقم (4.7) يوضح نتيجة التجربة السادسة

الخاصية	كل البيانات 14868	العنقود 0 1671	العنقود 1 928	العنقود 2 1999	العنقود 3 1713	العنقود 4 837	العنقود 5 1555	العنقود 6 1320	العنقود 7 1272	العنقود 8 1675	العنقود 9 1898
بناء المنزل	3.1	3.2	2.6	3.1	3.1	3	3	2.9	3.1	3.4	2.9
ملكية المسكن	6.9	7.4	2.3	7	7.9	2.2	8	7.9	7.6	6.8	8
الصرف الصحي	4.2	4.8	3	4.8	4.9	3.8	3.8	3.2	4.5	4.9	3.4
مصدر المياه الشرب	4.7	7.8	2.5	7.7	2	2.7	2.6	2.6	7.4	7.8	2.2
نوع الإضاءة	5	7.5	2.8	7.9	4.7	3.4	3.4	3.3	2.9	7.9	3.2
وقود الطهي	5	6.5	3.3	6.4	6.2	4.2	3.7	3.1	5	6.7	3.1
الصحة	7	7	7	7	7.1	7	7.1	7.1	7	7	7.1
نوع التعلم	2.3	2.2	2.3	2.1	2.9	2.4	1.9	1.7	1.8	3.3	1.8
عدد الغرف	3.5	1.8	4.2	4.7	3.5	2.5	1.8	4.7	4	4	3.6
نوع الاحتياجات	4.5	4.6	5.2	4.4	5.2	4.3	4	4.3	4.6	4.5	3.8
دخل الأسرة	2.7	2.9	2.8	2.7	3	2.8	2.6	2.4	2.6	3	2.4
عدد الأسرة	3	1.9	4.1	4.6	2.5	1.8	1.2	5	3.5	2.4	2.6
المجموع	46.8	61.4	28.6	67.3	54.4	27	25.9	36.4	53.7	66.4	25.5

4.8.1 تحليل النتائج:

العنقود 0 يمثل المجموعة الثانية وتحتوي على 1671 سجل وهي 11% من نسبة البيانات

العنقود 1 يمثل المجوعة الاولي وتحتوي على 928 سجل وهي 6 % من نسبة البيانات

العنقود 2 يمثل المجموعة الثالثة وتحتوي على 1999 سجل وهي 13% من نسبة البيانات

العنقود 3 يمثل المجوعة الثانية وتحتوي على 1713 سجل وهي 12% من نسبة البيانات

العنقود 4 يمثل المجوعة الاولي وتحتوي على 837 سجل وهي 6 % من نسبة البيانات

العنقود 5 يمثل المجوعة الاولي وتحتوي على 1555 سجل وهي 10% من نسبة البيانات

العنقود 6 يمثل المجوعة الثانية وتحتوي على 1320 سجل وهي 9% من نسبة البيانات

العنقود7 يمثل المجوعة الثانية وتحتوي على 1272 سجل وهي 9% من نسبة البيانات

العنقود8 يمثل المجوعة الثانية وتحتوي على 1675 سجل وهي 11% من نسبة البيانات

العنقود9 يمثل المجوعة الاولي وتحتوي على 1898 سجل وهي 13% من نسبة البيانات

4.8.2 ملخص التجربة:

في هذه التجربة تظهر لنا:

نجد انه تكررت نتائج التجربة السابقة من حيث ظهور المجموعات

ظهرت لنا العناقيد الثلاثة

الخلاصه والتوصيات

5.1 الخلاصة :

تناول البحث تحليل بيانات الحصر الشامل للفقراء الذي يعد الخطوة الاولي نحو صياغة استراتيجية لتخفيف ظاهرة الفقر. فالدخل المتدني لا يعتبر السمة الوحيدة للفقر فعادة ما يرتبط الفقر بسوء التغذية ومستويات التعليم المتدنية والظروف المعيشية المتواضعة وصعوبة الحصول علي الخدمات الاساسية مثل التعليم والصحة والمياه ومن ثم عملية توزيع الخدمات الاجتماعية في السودان لابد ان نرتكز علي الاعداد الفعلية للسكان الفقراء بالاضافة الي خصائص هؤلا السكان الذين يعيشون تحت خط الفقر.

الهدف من هذا البحث هو معرفة هل تصنيف فئات الفقر الذي تتبعه الأمانة العامة لديوان الزكاة مثالي ام ان هنالك تصنيف آخرا كثر ملائمة نتائج البحث اثبتت ان:

- معظم النتائج تظهر الفئات الاولي و الثانية.
- الفئه الثالثة الاقل ظهورا في النتائج.

1.2 التوصيات

من خلال تطبيق خوارزمية k-means علي بيانات الحصر الشامل للاسر الفقيرة نجد ان الطريقة المتبعة في هذا البحث افرزت نتائج توضح ان هناك تباين كبير في النتائح مما يستدعي إعادة تصنيف فئات الفقر الذي تتبعه الامانة العامة لديون الزكاة بزيادة عدد فئات الفقر الي اكثر من ثلاث فئات.

المراجع:

اولاً : المراجع باللغة العربية :

[3] المعهد العالي لعلوم الزكاة –احصاءات الفقراء، 2011م

[4] بسام علي،استخدام تنقيب البيانات للتنبؤ بظاهرة تسرب الطلبة، المجلة العراقية للعلوم الاحصائية، 2010م

[9]محمد مصطفي حجوز --تحسين خوارزمية K-means– مجلة جامعة البحث , 2014م

[10] د. اياد زعرور ،التنقيب في البيانات والتوقعات الاقتصادية

ثانياً : المراجع باللغة الإنجليزية:

[1] Two Crows Corporation, Introduction to Data Mining and Knowledge Discovery http://www.twocrows.com/intro-dm.pdf. 4/4/2015

[2] Data Mining: Techniques to Gain Insight Into Your Data Enterprise Systems Journal (December 1995):18-24, 64

[6] R. Bagherzadeh, ATMT Research Institute, Amirkabir University of Technology, Tehran, Iran.E-mail: Bagherzadeh_r@aut.ac.ir, *World Appl. Sci. J., 8 (8): 923-929, 2010*

[7] AGGARWAL, C.C., PROCOPIUC, C., WOLF, J.L., YU, P.S., and PARK, J.S. 1999a. *Fast algorithms for projected clustering*. In Proceedings of the ACM SIGMOD Conference,61-72, Philadelphia, PA

[8] R. Kirkby, WEKA Explorer User Guide for version 3-3-4, University of Waikato, 2002.

الملاحق:

بسم الله الرحمن الرحيم

الأمانة العامة لديوان الزكاة

الحصر الشامل للفقراء 2011

1/الولاية: ..... 2/ المحلية: ..... 3/ الوحدة الإدارية: .....

المدينة : ..... مجلس الحي : ..... المربع : ..... الحارة : ..... القرية : .....

م	جميع أفراد الأسرة						الأفراد 6 سنوات فاكثر
							الحالة التعليمية
رقم متسلسل	الاســـــم رباعـــي	صلة القرابة برب الأسرة	النوع: 1/ ذكر 2/ أنثي	العمر بالسنوات الكاملة	الديانة 1- مسلم 2- مسيحي 3- معتقدات 4-أخرى	هل يعاني للاسم مرض مزمنا 1. نعم 2. لا	1. أمي 2. يقرأ ويكتب 3. ابتدائي 4. ثانوي 5. جامعي 6. فوق الجامعي
	1	2	3	4	5	6	7
1		رب الاسرة					
2							
3							
4							
5							
6							
7							
8							
9							
10							

1

رقم متسلسل	الأفراد من 12 فاكثر	الأفراد 10 سنوات فاكثر					
	الحالة الزواجية : 1.أعزب 2.متزوج 3.مطلق 4.أرمل 5.مهجور/منفصل	إذا لم تكن تعمل لماذا؟ 1.سبب المرض 2. سبب العجز 3. لم أجد عملاً 4. لم أبحث عن عمل 5. لا ينطبق 6.أخرى تحدد	مصادر هل لديك دخل أخري 1.نعم 2.لا	الدخل الشهري (بالجنيه السوداني)	إذا كانت الإجابة (نعم) في الحقل (8) حدد المهنة 1.عامل حكومي 2.عامل غير حكومي 3.موظف حكومي 4.موظف غير حكومي 5. حرفي 6. بائع 7. تاجر 8. مزراع 9. أخري حدد	العلاقة بقوة العمل هل تعمل حاليا 1.نعم 2.لا
	13	12	11	10	9	8
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

2

(14) دخل الأسرة من النشاط الزراعي	الدخل بالجنيه للشهر الماضي	الدخل بالجنيه للعام الماضي
1. الدخل من المحاصيل		
2. منتجات البساتين		
3.الحيوانات		
4.منتجات الحيوان		
5.الدواجن ومنتجات الدواجن		
6.الأسماك ومنتجات الأسماك		
7 . منتجات الغابات		
8 . أنشطة زراعية أخري		

(15) دخل الأسرة من الأنشطة الغير زراعية	الدخل بالجنيه للشهر الماضي	الدخل بالجنيه بالجنيه للعام الماضي
1- الدخل من المرتبات والأجور		
2- الأنشطة الصناعية		
3- التجارة		
4- إيجار العقارات		
5- إيجارات أخري		
6- النقل والاتصالات		
7- التحويلات من الخارج		
8-التحويلات من الداخل		
9- الإعانات		
10- مصادر أخرى		

(3)المستوى المعيشي للأسرة

مستوى الخدمات المقدمة

م	السؤال	الإجـابة	الرقم	الأسئلة	الإجـــابة
16	كم تستهلك الأسرة في المتوسط في اليوم الواحد	☐☐☐ جنيه سوداني	17	في رأيك هل يكفي دخل الأسرة الحالي لتغطية حاجاتها	1- نعم 2- لا
18	ماذا كنت تفعل عندما لا تجد المال اللازم لشراء الطعام والشراب	1-استدين 2- انازل عن بعض الوجبات 3- الجأ الى انواع أقل تكلفة 4- ابحث عن أي عمل 5- أبيع بعض الأشياء 6- الجأ الى آخرين للحصول على موارد الأكل والشراب 7- لا ينطبق (لم يحدث لي مثل هذا لموقف) 8- أخري (حدد) :	19	الأسئلة : ما رأيك في الفقر ؟	الأجوبة : 1. الفقر ليس عيباً 2. الأرزاق بيد الله 3. الفقر مصيبة 4. يجب أن يجتهد الإنسان حتى لا يكون فقيراً 5. لا ادري 6. أخري (حدد) .....
20	ماذا كنت تفعل اذا لم تجد المال اللازم لشراء أشياء ضرورية غير الطعام والشراب	1- ليس لدي أشياء ضرورية غير الطعام والشراب 2- أستدين 3- انازل عن بعض الوجبات 4- الجأ الى انواع أقل تكلفة 5- ابحث عن أي عمل 6- أبيع بعض الأشياء 7- الجأ الي آخرين للحصول عليها 8- لا ينطبق(لم يحدث لي مثل هذا الموقف) 9- أخرى حدد	21	هل تعتبر نفسك فقيراً	1- نعم 2- لا 3- لا ادري

م	السؤال	الإجابة
22	هل لديك أبناء تركوا الدراسة	1- نعم 2- لا 3- ليس لدي أبناء
23	إذا كانت الإجابة ب (نعم) للسؤال (22) أذكر الأسباب	1. عدم رغبة الابناء في مواصلة تعليمهم 2. لعدم القدرة على توفير متطلبات تعليمهم 3. لظروف مرت بالأسرة 4. أسباب ثقافية 5. لا ينطبق 6. أخري (حدد) .....

5

خبرات المبحوث مع مؤسسات الرعاية الاجتماعية

م	الأسئلة	الأجوبة	الرقم	الأسئلة	الإجابة
24	هل سبق أن تلقيت أي خدمة من أي مؤسسة	1- نعم 2- لا	27	ما نوع الخدمة/الخدمات التي حصلت عليها	1. مشروع إنتاج حيواني/زراعي 2. وسيلة إنتاج 3. مشروع صناعي/حرفي 4. خدمات 5. رأس مال تجاري 6. إعانة مالية 7. - علاج 8. غرم 9. رسوم دراسية 10.أخرى(حدد).....
25	إذا كانت الإجابة ب(نعم) للسؤال رقم 24هل كانت الخدمة من ديوان الزكاة أم من مؤسسة أخرى	1.من ديوان الزكاة 2.من مؤسسة اخرى 3.من كلا الاثنين 4. لا ينطبق	28	كيف تحصلت على الخدمة	1. استعنت بأحد معارفي بالديوان 2. سعيت لو حدي 3. بواسطة اللجنة الشعبية 4. بواسطة اللجنة القاعدية بالحي/ القرية 5. لم افعل شيئاً 6. أخرى(حدد).....
26	حدد عدد مرات الحصول على الخدمة	1- مرة واحدة 2- مرتان 3- ثلاثة مرات فاكثر	29	ما هي الأشياء التي لم تعجبك في هذه الخدمة	1- المبلغ ما كان كافي 2- المشروع لم يكن مدروساً 3- نوع المواد لم تكن جيدة 4- معالجة مستخدمي الديوان لم تعجبني 5- أخري (حدد) .....

6

خبرات المبحوث مع مؤسسات الرعاية الاجتماعية

م	الأسئلة	الأجوبة	الرقم	الأسئلة	الإجابة
30	من الذي اختار لك نوع الخدمة التي حصلت عليها	1- بنفسه 2- ناس الزكاة 3- أسرتي 4- الأصدقاء 5- ناس اللجنة في الحي 6- أخرى (حدد)	37	هل كانت تتم متابعة للمشروع بعد تنفيذه من قبل الديوان	1- نعم 2- لا 3- لاينطبق
31	إذا كنت قد حصلت على خدمة فهل تعتقد أنك استفدت منها ؟	1- نعم 2- لا 3- لاينطبق	38	لو أتيحت لك فرصة للحصول على خدمة من ديوان الزكاة فماذا تفضل أن تكون	1- إعانة مالية 2- دعم عيني 3- مشروع 4- لا أرغب في الحصول على خدمة من الديوان 5- أخرى (حدد) .....
32	إذا كانت الإجابة بنعم للسؤال 34 اذكر بعضاً من أوجه الاستفادة	1- دخلي الحسن 2- سددت ديوني /حليت مشكلة 3- اطعمت حاجات ما كنت اعرفها 4- اكتسبت علاقات جديدة 5- بقيت الفدر اسدد رسوم المدرسة 6- نوع الاكل في البيت صار أحسن 7- عملت تحسينات في البيت 8- لاينطبق 9 أخرى (حدد)	39	إذا كان ما تفضله مشروعاً فأيهما تفضل : المشروع الفردي أم الجماعي ؟	1. مشروع فردي 2. مشروع جماعي 3. أخرى حدد
33	إذا كانت الإجابة ب لا للسؤال 34 فلماذا في رأيك لم تستفد	1- المبلغ غير كافي 2- تصميم المشروع ما كان جيد 3- لم تكن هناك دراسة جدوى 4- لم تكن هناك متابعة 5- لاينطبق 6- أخرى حدد	40	هل سبق ان قابلت لجنة للزكاة بالحي	1- نعم 2- لا
34	إذا لم تحصل حالياً على أي نوع من الخدمة من ديوان الزكاة فما السبب في رأيك	1- ليس لدي واسطة 2- لا توجد خدمات للديوان بمنطقتنا 3- لا أدري 4- أخرى حدد	41	هل تعتقد ان طريقة عمل اللجنة بحيكم مناسب	1- نعم 2- لا 3- لاينطبق
35	هل تعرف أسر بمنطقتكم هي مستحقة للزكاة ولم تحصل عليها	1- نعم 2- لا	42	هل أقام الديوان أي مشروعات بحيكم	1- نعم 2- لا
36	إذا كانت الإجابة للسؤال 38 ب نعم في رأيك لماذا ؟	1- أصلاً لا توجد خدمات زكاة بالمنطقة 2- ليس لدي هذه الاسرة معارف بديوان الزكاة 3- بسبب اسلوب ناس اللجان 4- موارد الديوان لا تكفي 5- لاينطبق 6- أخرى حدد.....	43	إذا كانت الإجابة للسؤال 45 ب نعم هل تعتقد بأن المشروعات ناجحة ؟	1- نعم 2- لا
			44	في رأيك ممكن مشكلة الفقر تعالج ؟	1- نعم 2- لا 3- لا أدري

خصائص المسكن :

عدد غرف النوم بالمسكن	نوع بناء المنزل	نوع ملكية المسكن	نوع المرحاض	مصدر مياه الشرب	نوع الإضاءة	وقود الطهي
45	46	47	48	49	50	51
	1. جالوص 2. طوب محروق 3. جالوص وطوب 4. اسمنت 5. طوب واسمنت 6. قش وعيدان 7. أخرى حدد	1. ملك 2. إيجار 3. حيازة 4. تركة 5. مع أسرتي 6. مع أسرة شريكي الزوجي 7. أخرى حدد.....	1. حفرة 2. مساس 3. سايفون 4. جردل 5. أخرى 6. لايوجد	1. ماسورة داخل بالمنزل 2. ماسورة عامة 3. كارو مياه 4. ماسورة الجيران 5. أخرى حدد	1. كهرباء عامة 2. مولد بالمنزل 3. مولد بالحي 4. شمع 5. لمبة كيروسين 6. اخري حدد..... 7. - لايوجد	1. كهرباء 2. غاز 3. كيروسين 4. فحم 5. حطب 6. أخرى حدد

ملاها .....

راجعها: .....

المشرف : .....