جامعة السودان للعلوم والتكنولوجيا

كلية الدراسات العليا

قسم علوم الحاسوب وتقانة المعلومات

أفضل المعايير للملابس السودانية باستخدام تقنيات العنقدة

The Best Stander For Sudanese Clothes Using Clustering Techniques

بحث تكميلي مقدم لنيل درجة الماجستير في علوم الحاسوب

إعداد الطالب

نسرين حسن مدني محمد أحمد

إشراف الدكتور

محمد الحافظ مصطفي موسي

2014م

الآية

اقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ (1) خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ (2) اقْرَأْ وَرَبُّكَ الْأَكْرَمُ (3) الَّذِي عَلَّمَ بِالْقَلَمِ (4) عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ (5)

صدق الله العظيم

سورة العلق، (الآيات ١ – ٥)

الإهداء

أهدي ثواب هذا العمل المتواضع الذي احمد الله كثيراً على اكتمال نعتمه عليّ بأن وفقني في إكماله.

وإلى أمي التي وصاني الرسول (ص) عليها وساعدتني كثيراً

إلى أبي الذي بذل جهداً مقدراً لكي أرتقي إلى هذه المرحلة وما زال يدعو لي بالمزيد.

إلى أخوتي وهم نعم الساعد والسند

إلى كل من زودني معرفة وعلمني علماً

الباحثة

ت

الشكر والتقدير

أشكر والدي / ***حسن مدني*** رحمه الله وأسكنه فسيح جناته مع الصديقين والشهداء على أنك من رسخ بأعماق ذاتي التي هي زادي الأمثل لأكمل بها مسيرة قافلتي رأساً شامخاً أعتز بأني ممن تربى على يديك .

وأشكر أمي الحبيبة / ***عازه علي احمد*** ذات القلب الحنون التي دعمتني بدعواتها الصادقة و خفّفت عني الجهد والتعب جعل الله ما قامت به في ميزان حسناتها وأمدَّ في عمرها امي الحنونة .

ويسرني بأن أخص بالشكر والعرفان بالجميل : زوجي / ***ياسر عثمان*** على دعمه المعنوي ومساعدته لي في إكمال دراستي حيث كان خير عون لي طيلة حياتي الدراسية من تشجيع ودعاء وصبر وعطاء فجزاه الله عني خير الجزاء .

وكذلك أتقدم واخص بالشكر والتقدير رمز التواضع والعطاء ، مثال الأمل والتفاؤل كما أَدِينُ بعظيم الفضل والشكر والعرفان بعد الله سبحانه وتعالى في إنجاز هذا البحث وإخراجه بالصورة المرجوة ؛ إلى المشرف على الرسالة : الدكتور / ***محمد الحافظ مصطفي*** ، الذي منحني الكثير من وقته ، وجهده ، وتوجيهاته ، وإرشاداته ، وآرائه القيمة . ومدَّ يد العون لي دون ضجر للسير قدماً بالدراسة نحو الأفضل سائلة المولى القدير أن يجزيه عني خير الجزاء حفظه الله من كل سوء وجعله في ميزان حسناته إن شاء الله .

الشكر خالصه لجامعة السودان للعلوم والتكنولوجيا التي قدمت ومازالت تقدم العلم لطلابها.

وختاماً أسال الله العلي القدير أن يكون هذا العمل خالصاً لوجه ، وأن يجعله علماً نافعاً، ويسهّل لي به طريقاً إلى الجنة.

المستخلص

- هناك أنظمة عديدة ومختلفة لقياس الملابس في جميع أنحاء العالم.
- نسبة لإختلاف أحجام الشعوب تحاول هذه الأنظمة وضع معايير تغطي الغالبية العظمى من الأحجام بطريقة متوازنة .
- تنقيب البيانات هو العلم الذي يهتم بالبحث عن معلومات قيمة ومجهولة وسط زخم كبير من البيانات بعدة طرق ، يهدف هذا البحث لإستخدام العنقدة للوصول للنظام المناسب لمعايير الملابس النسائية السودانية .
- استخدمنا في هذا البحث خوارزمية k-means وهي من خوارزميات العنقدة الواسعة الاستخدام وذلك بإستخدام مجموعة بيانات جمعت من عدد من الخياطين السودانيين أجريت التجارب عليها للحصول علي أفضل عنقدة لهذه البيانات ومقارنتها بالمعايير الألمانية
- أظهرت معلومات جديدة وقيمة ، اختتم هذا البحث بعدد من التوصيات التي يمكن أن تساعد في إظهار معلومات جديدة وذلك بإضافة بيانات جديدة أو استخدام خوارزميات عنقدة آخري غير خوارزمية العنقدة المستخدمة في هذا البحث

Abstract

There are many different systems to measure clothes all over the world. These systems are trying to develop standards covering the vast majority of sizes in a balanced manner. Data Mining is the science that deals with the search for valuable information in a considerable huge data in several ways. This research aims to use clustering to find appropriate standard for Sudan female clothes sizes.

k-means algorithm -one of the widely used clustering algorithms- were used for this research. Sudanese female clothes data were collected from a number of tailors. Many experiments were conducted to obtain the best clustering of these data and compared Palmair European standard.

The results showed a new and valuable information , which lead to a number of recommendations that can help in designing Sudanese female clothes standard sizes. Better results and new more information can be found by adding more data or by using another clustering algorithms.

فهرس المحتويات

الموضوع		رقم الصفحة
الآية		أ
الإهداء		ب
الشكر والتقدير		ج
المستخلص باللغة العربية		د
المستخلص باللغة الإنجليزية		هـ
فهرس المحتويات		و
فهرس الجداول		ط
فهرس الأشكال		ح
الباب الأول: الإطار العام		
1-1	تمهيد	1
2-1	مشكلة البحث	1
3-1	أهداف البحث	1
4-1	منهجية البحث	1
5-1	هيكلية البحث	2
الباب الثاني: التنقيب في البيانات		
1	تعريف	3
2	وسائل التنقيب	3
3	أهم طرق التنقيب	4
2-1	قاعدة الارتباط	4
2-2	التصنيف	4
3-2	العنقدة	4
1-3-2	خوارزمية K-means	5
2-3-2	تطبيقات خوارزمية K-means	6
3-3-2	مزايا خوارزمية K-means	6
4-3-2	مساوي خوارزمية K-means	6
الباب الثالث: وصف البيانات والتجارب		

1	مدخل	7
2	وصف البيانات	7
2-1	جمع البيانات	7
2-2	تجهيز البيانات	7
2-3	مشاكل البيانات وكيفية تنقيتها	7
3	الأدوات والبرامج المستخدمة في تنقيب البيانات	8
الباب الرابع: النتائج ومناقشتها		
1	مقدمة	9
2	التجربة الأولى	12
3	التجربة الثانية	15
4	التجربة الثالثة	18
5	مناقشة النتائج	20
الباب الخامس: الخاتمة والتوصيات		
الخاتمة		21
التوصيات		22
المراجع		23
الملاحق		24-31

فهرس الجداول

رقم الجدول	اسم الجدول	رقم الصفحة
1	وصف البيانات	7
2	تقسيم العناقيد إلى ستة	12
3	جدول الصدر	13
4	جدول الخصر	14
5	تقسيم العناقيد إلى سبعة	15
6	جدول الصدر	16
7	جدول الخصر	17
8	تقسيم العناقيد إلى ثمانية	18
9	جدول الصدر	19
10	جدول الخصر	20

فهرس الأشكال

رقم الشكل	الموضوع	رقم الصفحة
شكل رقم (1)	خوارزمية الـ K-Means	6
شكل رقم (2)	برنامج (Weka)	9
شكل رقم (3)	البيانات داخل (Weka)	10
شكل رقم (4)	الخوارزمية داخل (Weka)	11

الباب الأول

المقدمة

(Introduction)

1.1 تمهيد

- مع كثرة البيانات الموجودة والمخزنة في ما يسمى بقواعد البيانات (Database) أصبحت موضع
- تساؤل عديد من الباحثين للاستفادة منها ، ومع زيادة انتشار مستودعات التخزين الضخمة ما يدعي (Data warehouses) ، جاءت فكرة الكشف والتنقيب علي هذه البيانات بطرق ذكية للمساعدة في حل المشاكل واتخاذ القرارات ، تنقيب البيانات عملية تحليلية تمزج بين علم الذكاء الاصطناعي وتعلم الآلة وقواعد البيانات وتعتبر خطوة من خطوات استكشاف المعرفة من قواعد البيانات تبين كيفية الاستفادة من عدد ضخم من قواعد البيانات وذلك بانتزاع واستخلاص أنماط مفيدة تمكننا من التنبيه والمساعدة في إتخاذ القرارات ،
- لقد أصبح اهتمام العديد من نخبة المطورين الي الاستفادة من السجلات ذات الأهمية مما ساعد في تطوير وتحسين بعض المشاريع ذات الأحجام الهائلة من البيانات الي استكشاف حلول ومعارف جديدة مفيدة لهم في المستقبل . **{1}**

1.2 مشكلة البحث

- لا توجد معايير للألبسة السودانية (standard) عند الخياطين والتي يمكن الاستفادة منها لمعرفة المعاير التي تناسب السودانيين ، ومن هنا يحاول البحث أن يستخدم طريقة تحليل العنقدة او التجمعات للبيانات (clusters) لتحقيق هذا الغرض

1.3 أهداف البحث

- إلقاء الضوء على أهمية الملابس في عصر تكنولوجيا المعلومات
- دراسة مشكلة تراكم البيانات ومحاولة إيجاد الحل الأمثل لها
- تطبيق تقنيات استخلاص المعرفة علي بيانات نظام المقاييس للتوصل لأفضل معايير للملابس النسائية السودانية ، تم جمع البيانات من الخياطين .
- محاولة تحسين مستوى استخدام المعلومات من قبل الخياطين من اجل زيادة الاداء.

1.4 منهجية البحث

- يتبع البحث المنهج التطبيقي التحليلي.
- عنقدة البيانات بعد جمعها باستخدام خوارزمية k-means وهي احدي خوارزمية العنقدة

1.5 هيكلية البحث

- تم تقسيم هذا البحث الي عدد من الأبواب تناولت مفردات هذا البحث الاتي :
- الباب الأول تناول تمهيد ، مشكلة البحث ، أهداف البحث ، منهجية البحث وهيكل البحث.
- الباب الثاني تناول تعريف واهم طرق تنقيب البيانات (Data Mining) ، وتناول البحث الخوارزمية المستخدمة بنوع من التفصيل .
- الباب الثالث تحدث عن وصف البيانات والتجارب ثم الأدوات والبرامج المستخدمة في البحث . الباب الرابع تناول النتائج ومناقشتها والتجارب بنوع من التفصيل.
- الباب الخامس تناول الخاتمة والتوصيات التي خرج بها هذا البحث

الباب الثاني

التنقيب في البيـــانات

(Data Mining)

التنقيب في البيانات
(Data Mining)

1. تعريف

- يقصد بتنقيب البيانات البحث عن المعلومات ذات العلاقة ببعضها البعض التي يجمعها خصائص وسمات مشتركة ويربطها وحدة الموضوع أو التخصص، ويتم البحث عن هذه المعلومات ذات العلاقة بين كم كبير جداً من المعلومات التي لا تربطها علاقة أو وحدة موضوعية واستخلاصها ثم تقديمها لمتخذ القرار.
- يمر تعدين البيانات بعدد من المراحل تبدأ من تنقية البيانات، ثم توحيد البيانات، واختبار البيانات ذات العلاقة، ثم نقلها، وتصنيفها ثم تقييمها واستخراجها. هذه هي مراحل تعدين البيانات ولكل مرحلة تفاصيلها الدقيقة، وتتم مرحلة تلو الأخرى لتخرج في النهاية معلومات مفيدة ذات علاقة وقواسم وروابط مشتركة، لكي تقدم في النهاية إلى صناع القرار والجهات المستفيدة بحيث يتم توظيفها والاستفادة منها في مختلف المجالات العلمية والعملية **{1}** .
- التنقيب في البيانات هي عملية بحث محوسب ويدوي عن معرفة من البيانات دون فرضيات مسبقة عما يمكن أن تكون هذه المعرفة. كما ويعرف التنقيب في البيانات على أنه عملية تحليل كمية بيانات (عادة ما تكون كمية كبيرة) لإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة لصاحب البيانات. يطلق اسم "نماذج **{2}**
- هناك نوعان أساسيان للتنقيب في البيانات هما :التنقيب ألاستشرافي والتنقيب الوصفي.
- التنقيب ألاستشرافي ينتج عنه نموذج عن النظام الذي تصفه البيانات المستخدمة في التنقيب. أما التنقيب الوصفي فينتج عنه معلومات جديدة بناء على المعلومات الموجودة داخل البيانات المستخدمة في عملية التنقيب **{3}**

2. وسائل التنقيب في البيانات

اختيار الوسيلة المناسبة يعتمد على طبيعة البيانات تحت الدراسة وعلى حجمها يمكن إجراء عملية التنقيب في البيانات بالمقارنة مع سوق البيانات ومخزن البيانات **{4}**

3. أهم طرق تنقيب البيانات

1.2 قاعدة الارتباط Association Rule

انها عملية استكشاف وتحليل كميات كبيرة من البيانات باستخدام أساليب آلية او شبه آلية اعتماداً علي اكتشاف نماذج وقواعد ذات مغزى .

2.2 التصنيف Classification

تحليل مجموعة من البيانات ووضعها علي شكل أصناف أو أقسام ، وهناك عدد من الطرق التي يمكن استخدامها في تصنيف البيانات باستخدام الخوارزميات .

3.2 العنقدة Clustering

جمع البيانات هي عملية وضع البيانات في تجمعات متشابهه، وهي أحد فروع التنقيب عن البيانات. تقسم خوارزمية التجميع مجموعة بيانات إلى عدة تجمعات، حيث أن التشابه بين النقاط ضمن تجمع معين أكبر من التشابه بين نقطتين ضمن تجمعين مختلفين. فكره تجميع البيانات بسيطة في طبيعتها وقريبة جداً من الإنسان في طريقة تفكيره حيث أننا كلما تعاملنا مع كمية كبيرة من البيانات نميل إلى تلخيص الكم الهائل من البيانات إلى عدد قليل من المجموعات أو الفئات، وذلك من اجل تسهيل عمليه التحليل.

تستخدم خوارزميات التجميع على نطاق واسع ليس فقط لتنظيم وتصنيف البيانات ولكن لضغط البيانات وبناء نموذج ترتيب البيانات ، حيث أنه إذا كان بإمكاننا أن نجد تجمعات من البيانات، فانه بالإمكان بناء نموذج للمشكلة على أساس تلك التجمعات.

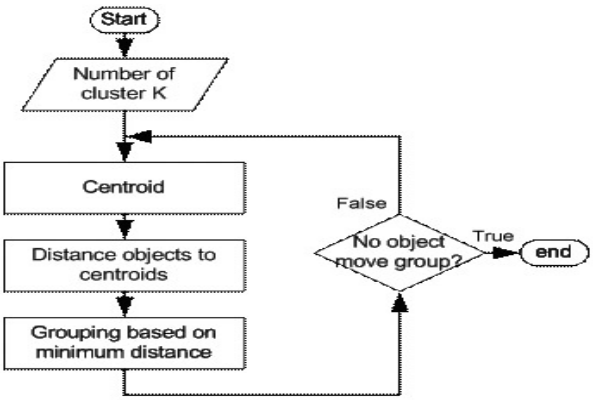
هي عملية تقسيم البيانات الي مجموعة من الأصناف اعتماداً علي اشتراكها بالخواص المتشابهة وان العنقدة هي تقسيم غير موجه للبيانات ، وهي عكس التصنيف كما أنها تساعد المستفيد علي فهم التركيب الطبيعي للمجموعات من البيانات .

، أما الطريقة الثانية فهي العنقدة الجزيئة وهي من أهم طرق العنقدة وهذه الطريقة تفترض ان كل العناصر الموجودة عبارة عن مجموعة واحدة ومن ثم تختار عنصرين بصورة عشوائية وترى قياس التشابه بين العنصرين واكبر العناصر تشابهاً هي التي تنقسم وهكذا تتوالى الانقسامات الي أن تصل لعدد الانقسامات المطلوبة ومن أشهر أمثلتها هي خوارزمية (k-means).**{5}**

1.3.2 خوارزمية K-means

- تستخدم خوارزميات التجميع على نطاق واسع ليس فقط لتنظيم وتصنيف البيانات ولكن لضغط البيانات وبناء نموذج ترتيب البيانات. حيث أنه إذا كان بإمكاننا أن نجد تجمعات من البيانات، فانه بالإمكان بناء نموذج للمشكلة على أساس تلك التجمعات.
- هناك العديد من الخوارزميات المستخدمة في عملية تجميع البيانات، وسندرس أبسط خوارزمية، ألا وهي خوارزمية k-means clustering تستخدم هذه الخوارزمية لتجميع عدة بيانات (أمثلة) اعتماداً على خصائصها إلى K تجمع ، وتتم عمليه التجميع من خلال تقليل المسافات بين البيانات ومركز التجمع (centroid cluster) وهي ايضاً واحدة من خوارزميات العنقدة التي تعمل علي تنقيب البيانات .
- استخدمت في هذا البحث إحدى طرق العنقدة المتمثلة بخوارزمية ال k-means مع المنطق في تصنيف البيانات واستخدمت هذه الخوارزمية في تكوين عناقيد للبيانات الموجودة التي تختلف في مكوناتها لاكتشاف بيانات ومكونات أخرى لم تظهر في البيانات الأصلية قبل تطبيق الخوارزمية .
- يوجد العديد من التقنيات التي يتم من خلالها التعامل مع البيانات ومن احد هذه التقنيات هي استخدام خوارزمية ال k-means .
- إذا أجريت العديد من الأبحاث في هذا المجال حيث استخدمت خوارزمية ال k-means في العنقدة عن طريق تحليل المكونات الرئيسية للبيانات الإحصائية .
- تستخدم هذه الخوارزمية لتجميع عدة بيانات (أمثلة) اعتماداً علي خصائصها الي k تجمع ، وتتم عملية التجميع من خلال تقليل المسافات بين البيانات ومركز التجمع (centriod cluster) .
- يعتمد أداء هذه الخوارزمية على المواقع الأولية لمراكز التجمعات (centroid) ، ومن المستحسن تنفيذ هذه الخوارزمية عدة مرات مع اختلاف المراكز في كل مرة عن المرات السابقة . **{6}**

يبين الشكل التالي خوارزمية K-Means Cluster



الشكل (1)

2.3.2 تطبيقات خوارزمية K-means Clustering؟

تعتبر خوارزمية K-Means Clustering من أبسط خوارزميات التجميع وأكثرها فعالية، وهناك الكثير من تطبيقاتها، أهمها:

- التعليم بدون إشراف unsupervised learning للشبكات العصبية.
- تمييز النماذج Pattern recognitions.
- التصنيف Classification.
- معالجة الصورة image processing.
- الرؤية بواسطة الآلة machine vision.
- الذكاء الصناعي Artificial intelligent.**{6}**

3.3.2 مزايا خوارزمية K-Means Clustering

1. ذات فعالية عالية.
2. سهلة التنفيذ .
3. تتعامل مع القيم المستمرة والقيم المتقطعة (الاسمية)

4.3.2 مساوئ خوارزمية K-Means Clustering:

1. تحديد عدد التجمعات K يتم عشوائياً وقبل معالجة الأمثلة.
2. حساسة للحالة الأولية. يؤدي اختيار حالات أولية متعددة إلى إعطاء نتائج مختلفة للتجمعات. ونتيجة لذلك قد تقع الخوارزمية في مشكلة النهاية المحلية.
3. شكل التجمع دائري لأنه يعتمد على حساب المسافة.**{6}**

الباب الثالث

وصف البيانات و التجارب

(Describe the data and experiences)

1. مدخل

يصف هذا الباب البيانات التي استخدمت في البحث ثم يشرح التجارب

2. وصف البيانات (Data set Description)

الجدول يوضح البيانات التي اجريت عليها التجارب و صفات البيانات

الجدول (1)

نوع البيانات	رقمية
الغرض	العنقدة
أسماء الصفات	الكتف ، الكم ، عرض الصدر ، الخصر ، الأرداف ، الطول
عدد السجلات	200
عدد الصفات	6
القيم المفقودة	لا توجد

1.2 جمع البيانات (data collection)-:

- إعتمد جمع البيانات بصورة أساسية على التدوين اليدوي للبيانات
- تم جمع البيانات من خياطين للسيدات حيث كانت البيانات في شكل (word) وتم تحويلها لجداول حسابية (Microsoft Excel File) .
- البيانات التي جمعت كان عددها (200) سجل لعدد (6) صفات من سيدات مختلفة أحجامهم ومقاساتهم ، والأعمار من (25–40) عام

2.2 تجهيز البيانات (Data set processing)

- تم تجهيز البيانات التي جمعت في الشكل الملائم لعملية التنقيب .
- وأخذت وقت طويل من زمن عملية التنقيب في تجميع وتجهيز البيانات ، البيانات التي تم جمعها غير مكتملة ومشوشة مما قد يخفي أنماط مفيدة ، بسبب تشويش البيانات

3.2 مشاكل البيانات وكيفية تنقيتها

كانت هنالك مشاكل مختلفة للبيانات فيما يلي بعض منها وكيفية حلها ::–

- البيانات الناقصة (missing value) نقول أن البيانات ناقصة أذا كان بها قيم غير موجودة في الحقل (null value) لان معظم عمليات التنقيب تفترض عدم وجود قيم ناقصة .
- البيانات الشاذة (outliers data)

- هي القيم التي تكون بعيدة المدى الطبيعي للبيانات
- نضع البيانات علي هيئة csv أي تحفظ علي هذا الشكل داخل weka
- مسح البيانات التي بها شوائب

3. الأدوات والبرامج المستخدمة في تنقيب البيانات

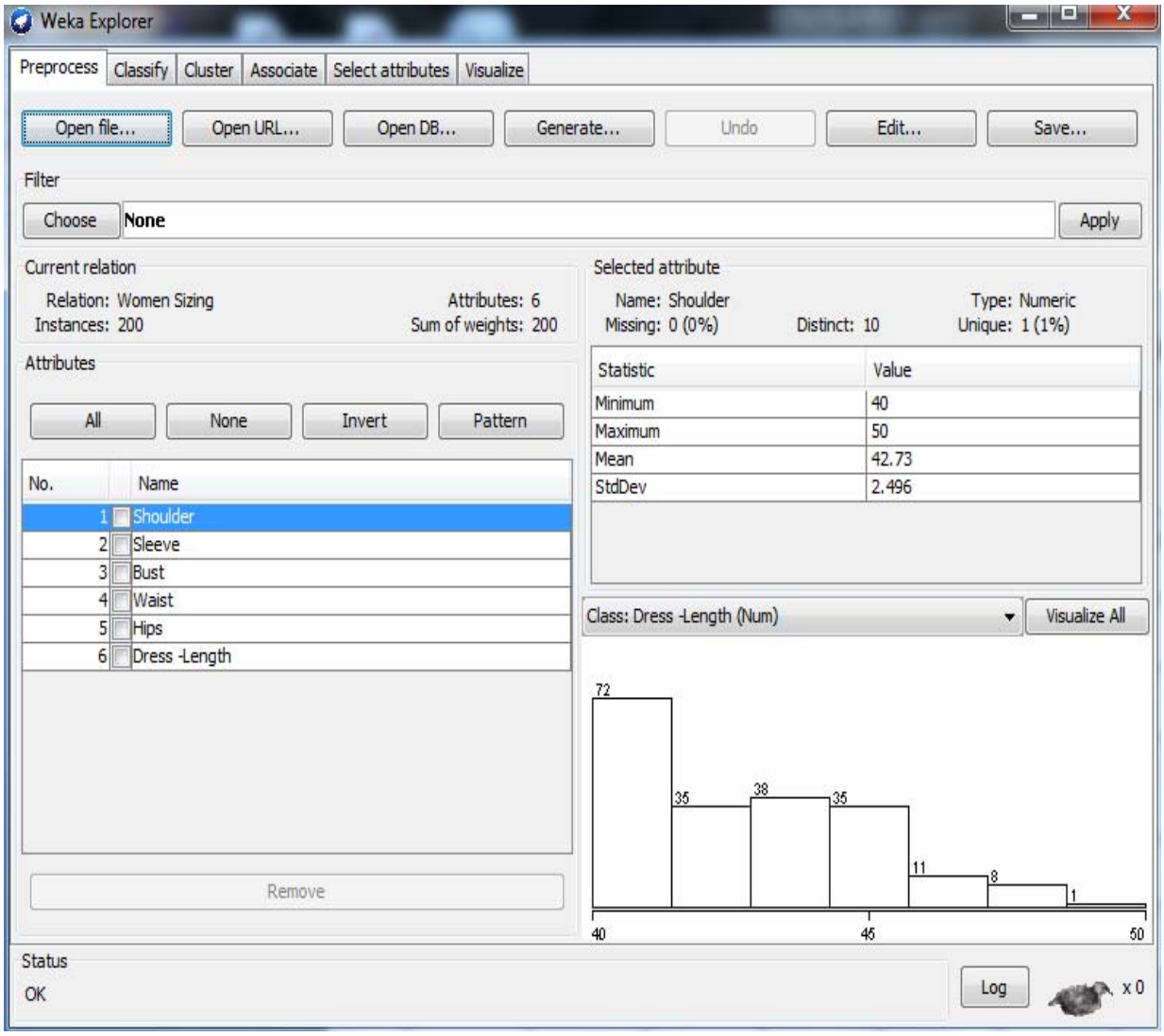
- إن الأدوات التي تستخدمها هذه التقنيات عديدة ومتنوعة ولكل واحدة منها دور ويخدم غرض معين، ومن هذه الأدوات (Weka)
- هناك العديد من المنتجات والأدوات والبرامج التي تقوم بعملية تنقيب البيانات والتي تنتج عن شركات كبيرة ومعروفة مثل (Oracle ، Spss ، IBM ، Microsoft)
- Weka هو عبارة عن برنامج خاص بخوارزميات التنقيب عن البيانات وتعلم الآلة
- كما يحتوي هذا البرنامج على أدوات قادرة على التعامل مع الأمور التالية pre-processing, classification, regression, clustering, association rules, and visualization
- يمكن إستخدامه من خلال الواجهة الرسومية مع واجهة المستخدم
  - هي واجهة و تنسيق البيانات
  - يستخدم ملفات نصية مسطحة لوصف البيانات
  - تحفظ ملفات البيانات في شكل ".arff" ، أو CSV
  - ويمكن أيضا أن يقرأ البيانات من URL أو من قاعدة بيانات SQL
  - تم تحميل برنامج weka من http://www.cs.waikato.ac.nz/ml/weka/
- تمت كتابة هذا البرنامج من قبل جامعة Waikato في نيوزيلاند ،وهو مكتوب بلغة الجافا ، ويعمل تحت رخصة GPL ، وهو يعمل تحت عدة بيئات (لينكس ، ويندوز ، ماك) {7}

1. مقدمة

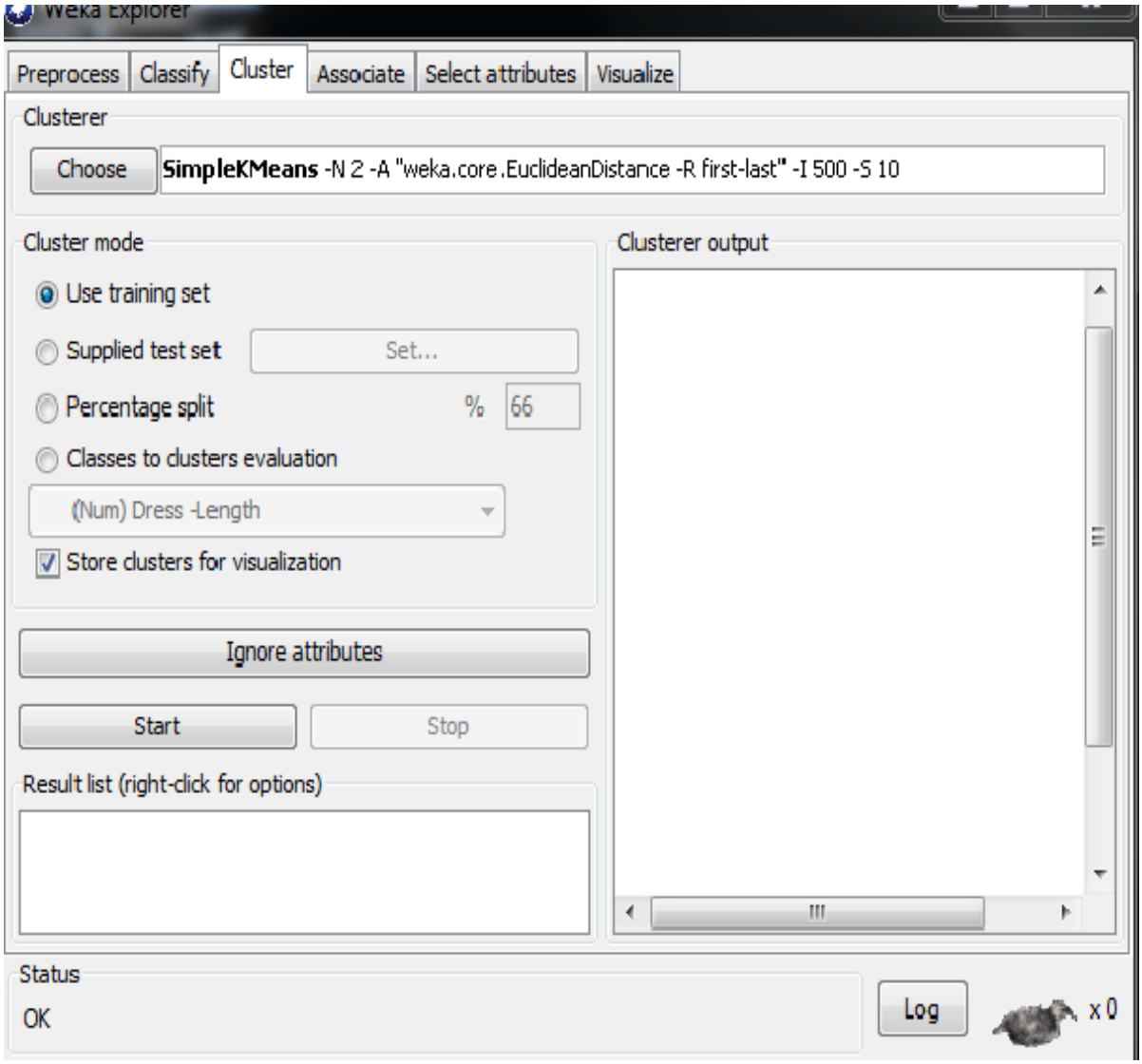
يصف الباب التجارب التي اجريت ونتائجها ويبدأ الباب بشرح مختصر لبرنامج الويكا (weka) ، اجريت ثلاثة تجارب ، التجربة الأولي للحصول علي ستة عناقيد ومقارنتها بالمعايير الألمانية ، التجربة الثانية للحصول علي سبعة عناقيد ومقارنتها بالمعايير الألمانية وكذلك الثالثة لنفس الغرض ، تم زيادة عدد العناقيد بواحد واثنين أي الي سبعة وثمانية علي المعايير الألمانية للحصول علي مزيد من المعلومات والتأكد هل العدد ستة هو المناسب للسودانيين ام أكثر .

Figure 1: Graphical User Interface Chooser

الشكل (2)


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Current relation
Relation: Women Sizing
Instances: 200
Attributes: 6
Sum of weights: 200
Selected attribute
Name: Shoulder
Missing: 0 (0%)
Distinct: 10
Type: Numeric
Unique: 1 (1%)
Attributes
All
None
Invert
Pattern
No.
Name
1 Shoulder
2 Sleeve
3 Bust
4 Waist
5 Hips
6 Dress -Length
Remove
Statistic
Value
Minimum 40
Maximum 50
Mean 42.73
StdDev 2.496
Class: Dress -Length (Num)
Visualize All
72
35
38
35
11
8
1
40
45
50
Status
OK
Log
x 0

الشكل (3) {7}


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
SimpleKMeans -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
%
66
Classes to clusters evaluation
(Num) Dress -Length
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
Clusterer output
Status
OK
Log
x 0

الشكل (4) {7}

الباب الرابع

النتائج ومناقشتها

(Results and discussion)

2. التجربة الأولي

جدول (2) : يوضح متوسطات العناقيد الستة الناتجة من التجربة الأولي {8}

الصفات	البيانات (200)	عنقود 0 (39)	عنقود 1 (39)	عنقود 2 (39)	عنقود 3 (29)	عنقود 4 (24)	عنقود 5 (30)
الكتف	42.7	40.5	45.1	40.8	45.1	45	40.6
الكم	58.4	58.7	57.7	58.9	59.6	58.3	57.0
عرض الصدر	99.9	113.1	99.2	98.3	115.1	84.7	83.3
الخصر	89.7	102.9	89.6	87.1	105.3	74.7	73
الأرداف	110.3	121.2	111.8	107.8	123.3	95.7	96.3
الطول	140.5	139.2	140.1	142.5	140.4	141.7	139.3

جدول (3) : هذا الجدول يقارن بين متوسطات الصدر للعناقيد الستة الناتجة من التجربة الأولي والمعيار العالمي

في المائة	العنقود	مجموعة البيانات	المعيار العالمي	المقياس
15%	عنقود 5	83.3	77.5	S
12%	عنقود 4	84.7	85.5	S
20%	عنقود 2	98.3	93.5	M
20%	عنقود 1	99.2	102.5	L
20%	عنقود 0	113.1	113.5	Xl
14%	عنقود 3	115.1	125.5	xxl

- بواسطة خوارزمية العنقدة تمت تقسيم العناقيد ومقارنة مجموعة البيانات مع المعايير العالمية الموجودة في الجدول أعلاه والنسب من مجموعة البيانات الكلية مثلا عنقود (5) يمثل 15 % من مجموعة البيانات ، وعنقود (1) يمثل 20% من البيانات .
- مجموعة البيانات وجدنا لها المتوسطات من برنامج الويكا باستخدام خوارزمية ال-k means
- عنقود (4) وعنقود (5) يشبهان S
- عنقود (0) وعنقود (3) يشابهان XL
- عنقود (1) يشبه L
- عنقود (2) يشبه M
- بينما العنقود (5) لا شبيه له في المقايس الألماني
- وكذلك العنقود (3) لا شبيه له في المقياس الألماني
- المعاير XXL لا يوجد عنقود يشابه

جدول (4) : هذا الجدول يقارن بين متوسطات الخصر للعناقيد الستة الناتجة من التجربة الأولي والمعيار العالمي

في المائة	العنقود	مجموعة البيانات	المعيار العالمي	المقياس
15%	عنقود 5	73	61	M
12%	عنقود 4	74.7	68.5	M
20%	عنقود 2	87.1	77	L
20%	عنقود 1	89.6	86	L
20%	عنقود 0	102.9	96.5	Xl
14%	عنقود 3	105.3	108.5	xxl

- بواسطة خوارزمية العنقدة تمت تقسيم العناقيد ومقارنة مجموعة البيانات مع المعايير العالمية الموجودة في الجدول أعلاه والنسب من مجموعة البيانات الكلية مثلا عنقود (5) يمثل 15 % من مجموعة البيانات ، وعنقود (1) يمثل 20% من البيانات .
- مجموعة البيانات وجدنا لها المتوسطات من برنامج الويكا باستخدام خوارزمية ال-k means
- عنقود (5) وعنقود (4) يشبهان M
- عنقود (1) وعنقود (2) يشبه L
- عنقود (0) يشبه XL
- عنقود (3) يشبه XXL
- المعايير XS, S لا يوجد عنقود يشابه

3. التجربة الثانية

جدول (5) : يوضح متوسطات العناقيد السبعة الناتجة من التجربة الثانية {8}

الصفات	البيانات (200)	عنقود0 (23)	عنقود1 (42)	عنقود 2 (33)	عنقود3 (29)	عنقود 4 (25)	عنقود 5 (25)	عنقود6 (23)
الكتف	42.7	40.7	45.4	40.9	45.0	43.9	40.3	40.7
الكم	58.4	58.3	57.9	59.3	59.8	57.8	57	58.5
عرض الصدر	99.9	114.1	98.2	97.2	0.115	82	85.1	109.3
الخصر	89.7	104.2	88.6	86.3	105.2	72.1	74.2	98.5
الأرداف	110.3	124.4	110.7	107.1	123.1	93.6	97.7	115.5
الطول	140.5	131.3	140.4	141.3	141.1	139.1	140.9	149.4

جدول (6) : هذا الجدول يقارن بين متوسطات الصدر للعناقيد السبعة الناتجة من التجربة الثانية والمعيار العالمي .

في المائة	العنقود	مجموعة البيانات	المعيار العالمي	المقياس
13%	عنقود 4	82	77.5	S
13%	عنقود 5	85.1	85.5	S
17%	عنقود 2	97.2	93.5	M
21%	عنقود 1	98.2	102.5	L
12%	عنقود 6	109.3	113.5	Xl
14%	عنقود 3	115	125.5	Xl
12%	عنقود 0	114.5	125.5	Xl

- الجدول أعلاه يوضح خوارزمية العنقدة لعرض الصدر تمت تقسيم العناقيد ومقارنة مجموعة البيانات مع المعايير العالمية الموجودة في الجدول أعلاه والنسب من مجموعة البيانات الكلية مثلا عنقود (5) يمثل 15 % من مجموعة البيانات ، وعنقود (1) يمثل 20% من البيانات .
- مجموعة البيانات وجدنا لها المتوسطات من برنامج الويكا باستخدام خوارزمية ال-k means
- عنقود (5) وعنقود (4) يشبهان S
- عنقود (1) يشبه L
- عنقود (2) يشبه M
- عنقود (3) وعنقود (0) وعنقود (6) يشبهان XL
- المعايير XS, XXL لا يوجد عنقود يشابه

جدول (7) : هذا الجدول يقارن بين متوسطات الخصر للعناقيد السبعة الناتجة من التجربة الثانية والمعيار العالمي .

المقياس	المعيار العالمي	مجموعة البيانات	العنقود	في المائة
Xs	77.5	72.1	عنقود 4	13%
XS	77.5	74.2	عنقود 5	13%
S	85.5	86.3	عنقود 2	17%
M	93.5	88.6	عنقود 1	17%
L	102.5	98.5	عنقود 6	12%
Xl	113.5	104.2	عنقود 0	12%
xxl	125.5	105.2	عنقود 3	14%

- في جدول الخصر أعلاه يوضح نتيجة هذه التجربة بواسطة خوارزمية العنقدة تمت تقسيم العناقيد ومقارنة مجموعة البيانات مع المعايير العالمية الموجودة في الجدول أعلاه والنسب من مجموعة البيانات الكلية مثلا عنقود (5) يمثل 13 % من مجموعة البيانات ، وعنقود (1) يمثل 17% من البيانات .
- مجموعة البيانات وجدنا لها المتوسطات من برنامج الويكا باستخدام خوارزمية ال-k means
- عنقود (5) وعنقود (4) يشبهان Xs
- عنقود (2) وعنقود (1) يشبهان S
- عنقود (3) وعنقود (0) يشبهان L
- المعايير M, XL XXL ،لا يوجد عنقود يشابه

4. التجربة الثالثة

جدول (8) : يوضح متوسطات العناقيد الثمانية الناتجة من التجربة الثالثة

الصفات	البيانات (200)	عنقود0 (42)	عنقود 1 (39)	عنقود 2 (34)	عنقود 3 (24)	عنقود 4 (23)	عنقود 5 (26)	عنقود6 (19)	عنقود7 (12)
الكتف	42.7	40.7	44.6	40.7	44.8	45.0	40.8	40.3	45.1
الكم	58.4	58.2	57.0	59.1	59.5	58.5	56.6	59.0	59.7
عرض الصدر	99.9	114.1	99.0	95.1	116.7	84.8	82.4	109.8	102.7
الخصر	89.7	104.5	89.7	84.2	106.8	74.9	72.1	98.1	92.5
الأرداف	110.3	124.4	112.8	105.6	125.3	95.6	95.6	114	111.8
الطول	140.5	132.0	145	143.2	143.5	141.6	138.1	147.8	131.7

جدول (9) : هذا الجدول يقارن بين متوسطات الصدر للعناقيد الثمانية الناتجة من التجربة الثالثة والمعيار العالمي .

في المائة	العناقيد	مجموعة البيانات	المعيار العالمي	المقياس
13%	عنقود 5	82.4	77.5	S
12%	عنقود 4	84.8	85.5	S
17%	عنقود 2	95.1	93.5	M
14%	عنقود 1	99.0	93.5	L
11%	عنقود 7	102.7	102.5	L
10%	عنقود 6	109.8	113.5	XL
12%	عنقود 0	114.1	113.5	Xl
12%	عنقود 3	116.7	125.5	xl2

- في جدول الصدر تم تقسيم العناقيد الي ثمانية ، ومقارنة العناقيد مع المعايير العالمية
- والعنقود 5 يمثل 13% من مجموعة البيانات اما العنقود 3 فيمثل 12% ايضا من مجموعة البيانات .
- المعيار العالمي 85.5 اقرب للمعيار 84.8 من مجموعة البيانات ويسمي S والمعيار العالمي 113.5 اقرب للمعيار 116.5 من مجموعة البيانات ويسمي XL وهكذا يتم مقارنة باقي المعايير الاخري بنفس الطريقة .
- المعيار 125.5 في المعايير العالمية لا يوجد معيار قريب منه من مجموعة البيانات
- عنقود (5) وعنقود (4) يشبهان S
- عنقود (3) وعنقود (0) وعنقود (6) يشبهان XL
- عنقود (7) وعنقود (1) يشبهان L
- عنقود (2) يشبه M
- المعايير XS, 2XL لا يوجد عنقود يشابه

جدول (10) : هذا الجدول يقارن بين متوسطات الخصر للعناقيد السبعة الناتجة من التجربة الثانية والمعيار العالمي .

في المائة	العناقيد	مجموعة البيانات	المعيار العالمي	المقياس
13%	عنقود 5	72.1	77.5	Xs
%21	عنقود 4	74.9	77.5	XS
17%	عنقود 2	84.2	85.5	S
%41	عنقود 1	89.7	85.5	S
%11	عنقود 7	92.5	93.5	M
%01	عنقود 6	98.1	102.5	L
12%	عنقود 0	104.5	113.5	XL
%21	عنقود 3	106.8	125.5	2XL

- جدول الخصر أعلاه تم تقسيم العناقيد الي ثمانية والعنقود 5 يمثل 13% من مجموعة البيانات اما العنقود 3 فيمثل 21% ايضا من مجموعة البيانات .
- المعيار العالمي 93.5 اقرب للمعيار 84.8 من مجموعة البيانات ويسمي S والمعيار العالمي 116.7 اقرب للمعيار 113.5 من مجموعة البيانات ويسمي XL وهكذا يتم مقارنة باقي المعايير الاخري بنفس الطريقة .
- المعيار 125.5 في المعايير العالمية لا يوجد معيار قريب منه من مجموعة البيانات
- عنقود (5) وعنقود (4) يشبهان Xsو عنقود (2) وعنقود (1) يشبهان S
- المعايير XL ، 2XL لا توجد عناقيد تشابهما
- المعايير M ، L لا توجد عناقيد تشابهما

4. مناقشة النتائج

- بعد إجراء التجارب علي ثلاثة مراحل إتضح أن المعايير الألمانية لا تناسب الأحجام السودانية وذلك للفروقات التالية
- ظهر عنقود بين المقاس M و L مما يشير أهمية وجود معيار في المدى يمكن أن نسميه

ML {8}

الباب الخامس

الخاتمة والتوصيات

(Conclusion and Future work)

الخاتمة

- في هذا البحث قامت الباحثة بجمع مجموعة بيانات من الخياطين السودانيين ، ثم درست برنامج الويكا وتعلمت كيف تستخدمه بكفاءة بعد ذلك قامت بإجراء عدة تجارب كانت نتائجها ، اتضح أن المعايير الاوربية لا تناسب الأحجام السودانية في بعض جزئيتها ، أهم نتيجة توصل لها البحث ظهور معيار جديد ML يناسب الأحجام السودانية .
- واخيراً كانت هنالك بعض التوصيات التي خرج بها البحث حيث تكون أساس عمل الباحثين من بعدى .

التوصيات

- من خلال النتائج التي ظهرت لنا أوصي بالاتي ::-
- زيادة كمية البيانات المستخدمة في عملية التنقيب للوصول لنتائج أدق
- إستخدام طرق عنقدة أخري مع مجموعة البيانات للحصول علي مزيد من المعلومات
- إجراء تجارب علي الحقول الأخرى (الكتف – الكم – الأرداف – والطول)

المراجع
References

{1} www.alriyadh.com/576881

{2} http://www.jilani.ly/eportal/teaching/cs808/Lecture5.pdf .

} ar.wikipedia.org/wiki/3{

} https://nal111.wordpress.com/4{

https://compuscienpedia.wordpress.com/. }5{

{6}www.alhasebat.net/vb/attachment.php?attachmentid=1905&d

{7} http://www.linuxac.org/forum/threads/31481-weka-

{8} en. Wikipedia.org/ wiki/clothing-sizes #women

الملاحـق

(APPNDEX)

مجموعة البيانات التي تم جمعها من الخياطين

الرقم	الطول	الأرداف	الوسط	الصدر	الكم	الكتف
.1	135	90	70	80	60	40
.2	140	80	63	74	55	42
.3	135	95	65	77	55	40
.4	160	63	66	76	60	44
.5	120	91	65	74	58	42
.6	130	97	71	80	56	42
.7	125	97	70	82	55	44
.8	110	40	80	90	58	47
.9	100	45	80	88	53	43
.10	145	40	82	92	55	45
.11	145	60	90	98	56	46
.12	140	40	80	90	62	42
.13	110	45	64	75	55	40
.14	130	45	88	98	60	40
.15	115	65	77	87	60	44
.16	122	45	100	107	58	47
.17	150	130	90	100	57	42
.18	145	100	72	80	60	42
.19	150	65	85	95	58	45
.20	140	45	71	80	58	47
.21	135	40	85	98	60	40
.22	145	50	80	90	58	44
.23	135	40	90	100	61	40
.24	145	46	86	97	58	44
.25	160	50	95	103	60	46
.26	135	55	78	89	57	40

45	58	90	80	80	120	.27
40	62	105	95	50	100	.28
44	62	108	100	50	130	.29
40	57	101	90	55	125	.30
48	60	97	87	60	130	.31
45	55	95	85	65	130	.32
40	60	96	85	60	150	.33
45	55	89	88	120	140	.34
45	62	95	85	75	150	.35
42	60	90	80	100	140	.36
47	60	91	81	95	140	.37
45	62	110	100	115	110	.38
45	61	98	88	100	120	.39
45	55	83	75	130	140	.40
44	60	92	82	95	130	.41
45	60	94	84	120	155	.42
44	61	93	82	100	145	.43
40	58	80	70	90	120	.44
48	60	90	80	115	150	.45
40	60	91	81	90	100	.46
44	57	78	70	65	150	.47
45	58	75	64	55	130	.48
42	58	81	70	110	125	.49
40	59	88	77	115	140	.50
40	57	77	65	65	105	.51
40	57	81	70	130	135	.52
40	62	120	110	90	140	.53
40	60	97	87	100	140	.54
48	58	81	70	100	142	.55
44	56	77	66	70	150	.56
40	59	88	77	85	110	.57
45	57	89	88	65	100	.58
44	60	97	85	69	110	.59
42	58	89	78	110	145	.60
42	59	80	71	116	135	.61
44	57	81	72	50	135	.62
45	59	85	74	100	150	.63
42	60	94	85	125	140	.64
40	56	80	70	125	144	.65

40	58	88	77	115	130	.66
42	60	90	80	120	165	.67
40	58	88	76	130	155	.68
40	56	121	110	125	150	.69
40	59	86	75	132	154	.70
42	57	80	71	82	142	.71
40	58	83	73	114	136	.72
40	60	90	80	125	130	.73
47	62	110	100	130	140	.74
40	58	88	78	140	155	.75
48	58	93	82	115	130	.76
42	62	115	104	120	140	.77
44	58	80	70	88	117	.78
44	60	105	95	170	140	.79
40	55	85	75	120	155	.80
44	56	118	105	130	100	.81
40	57	100	90	60	152	.82
40	60	98	88	85	132	.83
45	58	115	104	110	145	.84
44	59	120	110	134	150	.85
41	62	119	95	105	150	.86
40	57	103	98	142	145	.87
42	59	100	90	140	145	.88
40	57	98	85	75	139	.89
45	58	110	100	70	100	.90
41	61	109	97	155	160	.91
40	60	120	111	100	160	.92
44	62	113	103	70	165	.93
42	57	103	93	105	150	.94
42	58	99	88	110	130	.95
44	62	108	97	95	145	.96
44	57	85	75	80	140	.97
45	56	79	70	70	150	.98
40	57	88	76	90	160	.99
42	60	105	90	115	145	.100
40	55	90	81	110	135	.101
40	59	85	75	110	140	.102
44	55	96	86	110	155	.103
44	58	84	74	95	135	.104

40	57	95	85	120	148	.105
42	58	100	90	100	140	.106
40	56	90	81	75	100	.107
42	57	105	95	105	125	.108
40	59	112	100	98	152	.109
47	57	95	85	90	140	.110
40	58	85	74	100	150	.111
40	57	119	108	85	130	.112
45	62	122	110	120	100	.113
40	59	108	100	105	110	.114
44	61	111	102	85	140	.115
45	60	122	111	11	155	.116
42	59	97	88	95	160	.117
45	57	105	95	110	140	.118
44	58	115	110	54	110	.119
40	57	113	100	95	120	.120
40	60	110	103	100	150	.121
45	58	100	90	130	145	.122
45	57	110	102	120	130	.123
40	56	115	110	125	150	.124
40	60	105	95	100	160	.125
44	58	120	110	115	130	.126
42	58	100	90	115	135	.127
40	57	107	100	110	152	.128
40	60	110	103	98	155	.129
45	61	114	104	110	165	.130
40	62	94	84	110	160	.131
45	61	110	100	120	140	.132
50	58	120	110	70	160	.133
40	59	105	95	120	150	.134
40	60	103	93	100	140	.135
40	57	106	103	65	115	.136
45	59	88	79	60	120	.137
45	58	103	93	70	110	.138
40	60	95	85	110	120	.139
45	57	110	100	85	110	.140
40	58	116	106	45	135	.141
45	59	105	95	40	140	.142
40	57	98	88	60	135	.143

44	56	99	90	45	160	.144
42	58	108	97	70	120	.145
42	59	109	100	40	130	.146
44	58	97	88	65	125	.147
42	60	102	90	40	110	.148
43	58	104	94	45	100	.149
44	59	100	90	40	145	.150
42	57	98	78	60	145	.151
42	61	110	100	40	140	.152
40	60	104	95	45	110	.153
40	58	119	108	45	130	.154
44	59	105	95	65	115	.155
47	57	102	90	45	122	.156
42	58	100	90	130	150	.157
42	60	115	95	100	145	.158
45	59	103	93	65	150	.159
47	57	99	88	45	140	.160
40	60	109	90	40	135	.161
44	58	118	110	50	145	.162
40	55	119	110	40	135	.163
44	56	122	112	46	145	.164
48	60	89	79	50	160	.165
40	55	90	80	55	135	.166
42	57	81	72	80	120	.167
40	60	97	85	50	100	.168
41	56	120	110	50	130	.169
42	57	106	96	55	125	.170
48	59	100	90	60	130	.171
45	56	110	100	65	130	.172
40	58	116	105	60	150	.173
45	55	99	89	120	140	.174
45	62	120	110	75	150	.175
42	60	110	100	100	140	.176
47	60	108	98	95	140	.177
45	61	120	110	115	110	.178
45	59	105	95	100	120	.179
45	57	103	93	130	140	.180
44	60	118	107	95	130	.181
43	60	100	90	120	155	.182

44	60	111	100	100	145	.183
40	62	117	102	90	120	.184
48	61	120	110	115	150	.185
40	60	115	110	90	100	.186
44	58	107	93	65	150	.187
45	57	98	88	55	130	.188
42	59	102	90	110	125	.189
40	56	109	90	115	140	.190
40	58	119	108	65	105	.191
40	59	120	110	130	135	.192
40	61	123	112	90	140	.193
40	56	116	105	100	140	.194
48	58	119	109	100	142	.195
44	57	106	96	70	150	.196
40	56	104	94	85	110	.197
42	58	118	107	65	100	.198
44	60	120	110	69	110	.199
42	58	115	105	110	145	.200