|  |  |  |
| --- | --- | --- |
|  | دراسة تحليلية لأبرز الخصائص العامة لضيوف الرحمن وتمثيلها باستخدام التقنيات الحديثة للتنقيب في البيانات الكبيرة لدعم عملية اتخاذ القرار |  |
|  | محمود أحمد الجمل، مازن الشمرانيمعهد خادم الحرمين الشريفين لأبحاث الحج والعمرة **An Analytical Study of the Most Important Characteristics of the Guests of Al-Rahman and their Representation Using Modern Technologies to Mining Large Data to Support the Decision-Making Process**Mahmoud Elgamal, Mazin Alshamarani The Custodian of the Two Holy Mosques Institute for Hajj and Umrah Research |  |

**ملخص البحث (Abstract):**

تعتبر دراسة ديموغرافية الحجاج أحد أبرز الدراسات بمعهد خادم الحرمين الشريفين لأبحاث الحج والعمرة وذلك لوجود سجلات عدة حافلة بالبيانات في مجال خدمة ضيوف الرحمن، حيث تحتوي على الكثير من الخصائص والنماذج لأبرز اهتماماتهم وسلوكياتهم خلال رحلتهم لأداء المناسك، ففي كل عام يقوم فريق متخصص من الباحثين بدراستها في المعهد، حيث تشمل متغيرات الدراسة كافة المتغيرات الإسمية والترتيبية والكمية، بالإضافة إلى إيجاد العلاقات بين تلك المتغيرات وتحليلها لتعطي نتائج ذات أهمية للاستفادة منها في المجالات الاقتصادية والتخطيطية وإدارة الحشود واتخاذ القرارات وغيرها. ومع تطور تقنيات التنقيب في البيانات وزيادة مستوى دقتها أصبح بالإمكان استخلاص العلاقات المهمة بين تلك المتغيرات والتنبؤ بالاتجاهات المستقبلية والتي ستساعد متخذي القرار والمهتمين بشؤون الحجاج بشكل كبير وجذري. في هذا البحث قمنا ببناء قاعدة بيانات أولية عن الحجاج تشمل كافة المتغيرات موضع الدراسة (مرحلة إعداد البيانات)، ثم عملنا على تحويل هذه البيانات إلى نماذج مناسبة للتنقيب فيها (patterns)، ومن ثم استخراج المعلومات الهامة منها (Knowledge discovery) وأخيراً استخدام تقنيات التمثيل المرئي (Visualization) لعرض المخرجات على شكل معلومات تسهم بشكل أو آخر في دعم متخذي القرار.

The study of demographics of pilgrims is one of the most prominent studies in the Institute of the Custodian of the Two Holy Mosques for Hajj and Umrah Research, because of the existence of several records rich with data about in the services of guests of Alrahman, where they contain many characteristics and models of the most important interests and behaviors during their journey to perform rituals. Each year, a team of specialized researchers at the Institute, studies the variables of the study including all nominal, ordinal and quantitative variables, in addition to finding relationships between these variables and analyzing them to give important results to benefit in the economic and planning fields, crowd management, decision making and others. With the development of techniques of data mining and increase in the level of accuracy it is possible to draw important relationships between those variables and predict future trends, which will help decision makers and those interested in the affairs of pilgrims in a large and radical way.

In this research, we have built a preliminary database of pilgrims covering all the studied variables (data preparation stage). We then worked on converting this data into appropriate models for mining(patterns), extracting important information (Knowledge discovery), Visualization to display outputs in the form of information that contributes in one way or another to supporting decision makers.

**مقدمة:**

تعمل حكومة المملكة العربية السعودية على تجنيد أجهزة الدولة بكامل طاقاتها لخدمة ضيوف الرحمن وتيسير أعمال أداء المناسـك لحجاج بيت الله الحرام، ويعتبر معهـد خادم الحرمين الشريفين لأبحاث الحج والعمرة جهة ذات اختصاص وريادة في مجال الأبحاث المتعلقة بالحج والعمرة والزيارة ومرجعاً علمياً ومعلوماتياً أصيلاً.

وتأتي دراسة ديموغرافية الحجاج كأحد أهم وأقدم الدراسات في المعهد حيث بدأت عام 1408، والتي نحاول من خلالها وضع صورة كاملة وواضحة المعالم عن الأوضاع والملامح البشرية والسكانية والجغرافية والصحيــة والخدمية أمام متخذي القرارات ودعمها بحيث تكون مرجعاً علمياً شاملاً عن الحج يساهم في وضع السياسات المناسبة الحالية والمستقبلية التي تتعلق بالحج والعمرة، ويعتبر موسم الحج أكبر التجمعات الإسلامية على الإطلاق بل وعلى مستوى العالم، ولذلك عمد المعهد على القيام بدراسة استطلاعية سنوية لخصائص الحجاج الديموغرافية والاقتصادية والاجتماعية وذلك باستخدام أسلوب المعاينة الإحصائية في ظل غياب وصعوبة استخدام أسلوب الحصر الشامل لمثل هذه الخصائص لمجتمع الحجاج، وحيث إن لدينا هذا الكم الهائل للبيانات لعدة أعوام كان لابد من استخدام التقنيات الحديثة للتنقيب في هذه البيانات واستخلاص المعلومات الهامة والتي يصعب استخلاصها من أرقام وبيانات مختلفة. وعملية التنقيب في البيانات هي عملية إيجاد لأبرز الخصائص والأنماط والارتباطات ضمن مجموعات البيانات الكبيرة وذلك باستخدام تقنيات استخراج البيانات المختلفة مثل التعلم الآلي، والذكاء الاصطناعي (AI) والإحصاء، ويمكن استخدام تلك المعلومات عن الحجاج لزيادة الإيرادات وتذليل الصعوبات ونشر الوعي وغير ذلك. وفي هذه الورقة البحثية قمنا باستخدام البيانات المتاحة عن الحجاج في المعهد من عام 1432 وحتى عام 1438؛ أي بيانات 7 سنوات، وتحتوي بيانات كل سنة على حوالي 35 متغيراً اخترنا بعضاً من هذه المتغيرات للبحث والتي مثلت نقطة البداية لهذا العمل.

قمنا باستخدام عدة خوارزميات للتنقيب في البيانات مثل Classification, Ranking, Clustering, …etc. ثم تم توظيف تقنيات التمثيل المرئي (Visualization) لعرض النتائج، وتتناول الورقة مواضيع البحث على النسق التالي: في الجزء الثاني من الورقة استعراض عملية التنقيب في البيانات، وفي الجزء الثالث دراسة لبيانات الحجاج، وفي الجزء الرابع استنتاج النتائج والتحليلات.

**أهداف البحث:**

* التنقيب في البيانات الكبيرة للحجاج لاستخلاص العلاقات غير المباشرة الهامة.
* دراسة العلاقة بين المتغيرات الخاصة بالحجاج وإيجاد العلاقة بينها.
* دراسة أهمية المتغيرات المستقلة على المتغير الهدف وترتيبها حسب أهميتها.
* عرض النتائج بصورة يسهل فهمها وقراءتها باستخدام التقنيات الحديثة للتمثيل المرئي للبيانات.
* الاستفادة من نتائج الدراسة في النواحي الاقتصادية والتوعوية والتخطيطية للحجاج.

**التنقيب في البيانات Data mining**



شكل (1): تقاطع تخصصات متعددة كأدوات وأساسيات في علم التنقيب في البيانات

التنقيب في البيانات هو عملية اكتشاف الأنماط في مجموعات البيانات الكبيرة التي تتضمن طرقًا من أعمال التعلم الآلي (Machine learning) والإحصاء (Statistics) وأنظمة قواعد البيانات (Database systems) [4]، كما هما مبين في شكل رقم (1)، عليه فإن تنقيب البيانات هو مجال فرعي متعدد التخصصات من علوم الكمبيوتر مع هدف شامل لاستخراج المعلومات (بالأساليب الذكية) من مجموعة البيانات وتحويل المعلومات إلى بنية مفهومة لمزيد من الاستخدام [4,5,8,9,10]، تنقيب البيانات هو خطوة تحليل عملية "اكتشاف المعرفة في قواعد البيانات" [7]، بالإضافة إلى خطوة التحليل الأولي، فهي تتضمن أيضًا جوانب إدارة البيانات وجداول البيانات والمعالجة المسبقة للبيانات والاعتبارات النموذجية والاستدلالية ومقاييس الاهتمامات واعتبارات التعقيد وعمليات ما بعد المعالجة للهياكل المكتشفة والتمثيل المرئي (Visualization) والتحديث عبر الإنترنت (Online Updating) [4].

وعملية تنقيب البيانات تتكون من المراحل التالية:

تكامل البيانات (Data Integration)، اختيار البيانات (Data Selection)، تنظيف البيانات (Data Cleaning)، تحويل البيانات (Data Transformation)، التنقيب في البيانات (Data Mining)، تقييم الأنماط واكتشاف المعرفة (Pattern Evaluation and Knowledge Discovery)، وأخيرا اتخاذ القرارات (Decisions/Use of Discovered Knowledge).

**بيانات الحجاج**

تتكون بيانات الحجاج من استبانة سنوية توزع على عينة عشوائية من الحجاج في موسم الحج لاستطلاع آرائهم، حيث تتكون هذه الاستبانة من حوالي 35 سؤالا يمثل كلا منها متغيرا احصائيا (اسمي، ترتيبي، أو كمي) تمثل الخصائص الديموغرافية والاقتصادية وغيرها للحجاج، وبعد انتهاء جمع الاستبانات يتم حفظ وبناء قاعدة بيانات من إجابات الحجاج وحفظها للتحليل سنوياً.

وفي هذا البحث تم تجميع البيانات الخام لتلك الدراسات منذ عام 1432 هـ الى 1438 هـ، ثم وضع تلك البيانات في جداول اكسل، حيث تمثل الأعمدة في جداول الملف المتغيرات وتمثل الصفوف سجل البيانات.

**الخطوة الأولى (المعالجة الأولية للبيانات Data preprocessing):**

* + - * تم إجراء العديد من العمليات لمعالجة البيانات الخام التي تم انتقاؤها خلال المراحل السابقة، ضمت عمليات المعالجة: (أ) إقصاء السجلات التي تحوي بيانات غير كاملة، (ب) دمج بعض الحقول كدمج مدة البقاء داخل المدينة المنورة قبل وبعد أداء مناسك الحج، (ج) تحويل بعض الحقول من حقول ذات بيانات رقمية إلى حقول ذات فئات رقمية كتحويل بيانات العمر من قيمة 24 إلى فئة 20-30، ثم اختصارها أكثر لتكون <30 والتي تعني أكبر من 20 وأقل من 30، (ج) إجراء الإصلاحات الإملائية للبيانات، (د) توحيد محتوى البيانات لجميع السنوات كتوحيد قيمة "موظف حكومي" و "قطاع حكومي" واختصارها إلى ( Pub)، هـ) اختصار "ثلاثة أيام" إلى ( 3dys) وهكذا، و) اختيار عشوائي لبعض الجنسيات ذات العدد الكبير مثل: تركيا، باكستان، مصر، السعودية، السودان، الهند، اليمن.
			* حفظ البيانات في ملف بصيغة (csv) حتى يسهل التعامل معه.
			* تحديد المتغيرات الهامة من بين تلك المتغيرات لاستكشاف العلاقات غير المعروفة بينها.

**الخطوة الثانية (التنقيب في البيانات Data Mining):**

هناك العديد من البرامج المستخدمة في التنقيب في البيانات مثل: لغة R[11]، RapidMiner[12]، Weka[14] إلى غير ذلك من البرامج المتاحة وقد استخدمنا خوارزمية (Attribute Selected Classifier Algorithm[6] ) والتي تستخدم المصنف ( Naive Bayes Classifier[3]) لاستخلاص الخصائص ثم استخدام ([2,9] Gain Ratio Attribute Evaluation) لتقييم الخصائص ولترتيب تلك الخصائص حسب أهميتها تم استخدام خوارزمية( Ranker Algorithm[8]).

**الخطوة الثالثة تقييم الأنماط واكتشاف المعرفة (Pattern Evaluation and Knowledge Discovery):**

الاكتشاف المعرفي للبيانات هو عملية تحديد النماذج الصالحة والجديدة والمفيدة والسهلة الفهم في البيانات، ويتعلق الاكتشاف المعرفي بعملية استخراج المعرفة بالكامل، بما في ذلك كيفية تخزين البيانات والوصول إليها، وكيفية استخدام خوارزميات فعالة وقابلة للتطوير لتحليل مجموعات البيانات الضخمة، وكيفية تفسير النتائج وتصورها، وكيفية تصميم ودعم التفاعل بين الإنسان والآلة. والنماذج هي مجموعة من العناصر أو التراكيب المرتبطة بقوة (Strongly Correlated) وتمثل خصائص ذاتية هامة للبيانات، واكتشاف تلك النماذج لمجموعات البيانات الكبيرة يمثل نقطة التحول في التنقيب في البيانات.

**الخطوة الرابعة التمثيل المرئي (Visualization) للنتائج:**

تعتبر النتائج والأرقام في أي عملية صعبة الفهم والتصور العام، ولذا تأتي أهمية الــتمثيل المرئي **(Visualization)** لتوضح النتائج بأشكال سهلة الفهم والإدراك. وفي هذا البحث استخدمنا طريقة الــتمثيل الشجري (Tree Visualization [1,13]) لعرض النتائج وإظهارها بصورة سهلة الفهم، ويمثل شكل (2) الرسم التخطيطي للطريقة المستخدمة.

شكل (2): رسم تخطيطي لطريقة التنقيب في بيانات الحجاج

يوضح شكل (2) المنهجية المتبعة للتنقيب في بيانات الحجاج حيث:

1. قمنا بتجميع بيانات الحجاج من عام 1432~1438، ومنها تم تحديد الحقول الهامة لتكون موضع الدراسة، ثم عملية تنظيف البيانات ومعالجة القيم الشاذة أو المفقودة واستبعاد السجلات غير الكاملة، ثم وضع الاختصارات المناسبة مثلSaudi(Sa) وPublic employee (Pub) وتحويل المتغيرات الكمية إلى متغيرات ترتيبية ،...إلخ.
2. ثم بعد ذلك عملية التنقيب في البيانات (Data Mining) هي عملية أساسية حيث يتم تطبيق أساليب ذكية لاستخراج أنماط البيانات.
3. وتأتي عملية تقييم النماذج (Pattern evaluation) لتحديد النماذج المثيرة للاهتمام حقًا والتي تمثل المعرفة بناءً على مقاييس مثيرة للاهتمام.

وأخيراً (Knowledge Presentation) حيث يتم استخدام تقنيات التمثيل المرئي لتقديم النتائج والمعلومات المستخرجة إلى المستخدمين بصورة يسهل فهمها.

**النتائج والتحليلات:**

وفي هذا الجزء سنقوم بدراسة العلاقة بين عدة متغيرات واستخلاص أهم النتائج، ثم تمثيلها مرئياً:

1. **الجنسية والعمر والإنفاق(بالريال السعودي):**

تمت دراسة العلاقة بين الجنسية والعمر والانفاق بالريال السعودي خلال فترة الحج، وكانت النتائج كالآتي:

يتضح من جدول (1) أن أكثر الجنسيات إنفاقاً بالريال السعودي على الترتيب هي: السعودية، اليمن، مصر، وتركيا.

جدول (1): الجدول التكراري النسبي للجنسية ومعدل الإنفاق خلال فترة الحج بالريال السعودي.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| الجنسية/الإنفاق | 2000 > | 2000 - 4000 | 4000 - 6000 | 6000 - 8000 | 8000 - 10000 | >10000 |
| اليمن | 10.53 | 6.00 | 10.28 | 33.33 | 40.91 | 6.58 |
| مصر | 8.02 | 14.00 | 17.76 | 12.82 | 13.64 | 28.95 |
| السودان | 6.52 | 16.75 | 13.08 | 7.69 | 9.09 | 19.08 |
| السعودية | 17.15 | 4.25 | 28.97 | 38.46 | 22.73 | 2.63 |
| باكستان | 21.26 | 23.00 | 13.08 | 2.56 | 4.55 | 1.32 |
| الهند | 24.67 | 21.75 | 13.08 | 2.56 | 4.55 | 1.32 |
| تركيا | 11.84 | 14.25 | 3.74 | 2.56 | 4.55 | 40.13 |
| المجموع | 100% | 100% | 100% | 100% | 100% | 100% |

جدول (2): الجدول التكراري النسبي للعمر ومعدل الإنفاق خلال فترة الحج بالريال السعودي.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| العمر/الانفاق | 2000 > | 2000 - 4000  | 4000 - 6000  | 6000 - 8000  | 8000 - 10000  | >10000 |
| 50-40  | 36.14 | 33.83 | 35.85 | 42.11 | 33.33 | 39.07 |
| 40-30 | 30.02 | 26.82 | 24.53 | 13.16 | 14.29 | 14.57 |
| 30-20 | 11.65 | 8.02 | 11.32 | 18.42 | 14.29 | 3.31 |
| 60-50 | 15.76 | 18.55 | 16.04 | 10.53 | 19.05 | 35.10 |
|  >60  | 4.62 | 12.03 | 6.60 | 10.53 | 9.52 | 5.96 |
|  <20  | 1.81 | 0.75 | 5.66 | 5.26 | 9.52 | 1.99 |
| المجموع | 100% | 100% | 100% | 100% | 100% | 100% |

ومن جدول (2) يتضح أن أكثر الأعمار إنفاقا حسب الترتيب الفئوي للجدول: هي الفئة العمرية (40-50) ثم (30-40)، ثم الفئة (60-50)، ويبين شكل (3) تلك العلاقات باستخدام (Tree Visualization).



شكل (3): التمثيل المرئي للعلاقة بين الجنسية والعمر والإنفاق بالريال السعودي (<50 تعني 40-50 وهكذا).

1. **الجنسية والعمر وعدد مرات الحج:**

تمت دراسة العلاقة بين الجنسية والعمر وعدد مرات الحج السابقة، وكانت النتائج كالآتي حسب جدول (3) (بناء على ترتيب أكبر جنسيتين): ويلاحظ من الجدول أن جنسيات السعودية والهند وباكستان ومصر واليمن هم أكثر الجنسيات تكراراً للحج.

جدول (3): الجدول التكراري النسبي للجنسية والعمر وعدد مرات الحج.

|  |  |  |
| --- | --- | --- |
| عدد مرات الحج | الجنسية | العمر |
| لم يسبق له الحج | الهندباكستان | 40-5030-40 |
| مرة واحدة | الهندباكستان | 30-5020-30 |
| مرتين | السعوديةباكستان | 30-5050-60 |
| ثلاث مرات | السعوديةمصر | 30-5030-60 |
| أربع مرات | السعوديةمصر | 30-40>60 |
| خمس مرات | السعوديةاليمن | 30-5020-30 |
| ستة مرات | السعودية | 20-30 |
| سبع مرات | مصر | 40-50 |
| ثمان مرات | اليمن | 30-40 |
| عشر مرات | مصر | 40-50 |
| أكثر من عشر مرات | اليمنباكستان | 30-5020-30 |

ويوضح شكل (4) تلك العلاقات بصورة أوضح، حيث أن المتغير الهدف (عدد مرات الحج) كأصل الشجرة( Root) يتفرع منها عمر الحاج(’30> ‘ وتعني 20-30 أو ’50> ‘ وتعني 40-50 وهكذا )، ثم جنسية ذلك الحاج مختصرة مثلاً( Yemen -> YE, Pakistan -> PK)وهكذا. ولمعرفة جنسيات وأعمار الحجاج الذين أدوا الحج أكثر من 10 مرات نضغط على العدد ‘>10’، ثم على ورقة الأعمار فتظهر الأعمار، ومنها تنبثق الجنسيات كما بالشكل. ومن الصعوبة توضيح ذلك بالجداول حيث لابد من قراءة الأرقام وربطها ببعض وهذا صعب للغاية.



شكل (4): التمثيل المرئي للعلاقة بين الجنسية والعمر وعدد مرات الحج.

1. **الجنسية والدخل وزيارة المدينة المنورة:**

 تمت دراسة العلاقة بين الجنسية والدخل وعدد أيام البقاء في المدينة المنورة، وكانت النتائج كالآتي المدة الشائعة للبقاء بالمدينة المنورةهي 8 أيام، 7 أيام، ثم 5 أيام، وأن جنسيات تركيا والهند وباكستان والسودان واليمن قد مكثوا 8 أيام (جدول (4))، ويتضح من جدول (4) أيضا أن 68.21% من السعوديين "لا إجابة"، ومن جدول (5) يتضح أن أكثر فئات الدخل مكوثاً في المدينة المنورة، هم فئة الدخل "2000-4000"، تليهم فئة الدخل ">10000"، ثم فئة الدخل "4000-6000" ريال.

ويوضح شكل (5) التمثيل المرئي للعلاقة بين الجنسية والدخل وعدد أيام البقاء في المدينة المنورة.

جدول (4): الجدول التكراري النسبي للجنسية وعدد أيام البقاء في المدينة المنورة.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| اليمن | مصر | السودان | السعودية | باكستان | الهند | تركيا | عدد الأيام/الجنسية |
| 24.39 | 10.98 | 8.21 | 68.21 | 11.17 | 21.13 | 1.08 | لا إجابة |
| 2.44 | 5.28 | 0.97 | 1.43 | 5.59 | 4.9 | 0.72 | يوم |
| 3.41 | 3.66 | 1.45 | 13.21 | 4.47 | 2.84 | 1.44 | يومان |
| 18.54 | 6.1 | 2.9 | 4.64 | 1.96 | 1.55 | 7.91 | 3أيام |
| 17.07 | 15.45 | 1.93 | 1.79 | 0.28 | 1.55 | 9.35 | 4أيام |
| 9.76 | 21.95 | 16.91 | 3.57 | 5.03 | 2.32 | 11.87 | 5أيام  |
| 0.49 | 0.81 | 13.53 | 1.79 | 1.12 | 0.77 | 3.24 | 6أيام |
| 0.98 | 19.11 | 15.46 | 1.79 | 9.5 | 7.99 | 16.91 | 7أيام |
| 18.54 | 10.98 | 28.99 | 0.71 | 55.59 | 36.86 | 35.97 | 8أيام |
| 1.46 | 0.81 | 3.86 | 0.36 | 4.47 | 1.8 | 6.47 | 9أيام |
| 1.95 | 3.66 | 3.86 | 1.07 | 0.28 | 16.24 | 3.96 | 10أيام |
| 0.98 | 1.22 | 1.93 | 1.43 | 0.56 | 2.06 | 1.08 | أكثر من 10 أيام |
| 100% | 100% | 100% | 100% | 100% | 100% | 100% | المجموع |

جدول (5): الجدول التكراري النسبي للدخل وعدد أيام البقاء في المدينة المنورة.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| الدخل/عدد الأيام | لا إجابة | يوم | 2يوم | 3يوم | 4يوم | 5يوم | 6يوم | 7يوم | 8يوم | 9يوم | 10يوم | >10 |
| لا إجابة | 3.38 | 8 | 1.75 | 1.69 | 3.45 | 3.94 | 3.13 | 7.5 | 1.54 | 3.23 | 2.86 | 6.67 |
| 2000 - 4000 | 26.32 | 56 | 31.58 | 42.37 | 29.31 | 38.58 | 53.13 | 51.25 | 54.46 | 35.48 | 37.14 | 33.33 |
| 4000 - 6000 | 7.89 | 12 | 10.53 | 15.25 | 12.07 | 22.05 | 18.75 | 23.75 | 8.92 | 22.58 | 20 | 13.33 |
| 6000 - 8000 | 10.53 | 4 | 7.02 | 15.25 | 12.07 | 8.66 | 18.75 | 5 | 7.69 | 22.58 | 5.71 | 20 |
| 8000 - 10000 | 19.55 | 16 | 15.79 | 8.47 | 24.14 | 14.17 | 3.13 | 5 | 6.15 | 6.45 | 5.71 | 13.33 |
| >10000 | 32.33 | 4 | 33.33 | 16.95 | 18.97 | 12.6 | 3.13 | 7.5 | 21.23 | 9.68 | 28.57 | 13.33 |
| المجموع | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |



شكل (5): التمثيل المرئي للعلاقة بين الجنسية والدخل وعدد أيام البقاء في المدينة المنورة.

1. **الجنسية والعمر والوظيفة والدخل وعدد ليالي المبيت في منى:**

تمت دراسة العلاقة بين متغيرات الجنسية والعمر والوظيفة والدخل وعدد ليالي المبيت في منى كمتغير مستقل، فكانت النتائج كالآتي:

1. كانت الجنسية التركية هي أكثر الجنسيات مبيتا لليلة واحدة، والهند لليلتان، وباكستان لثلاث ليالي.
2. كانت أكثر المهن مبيتاً بمني هي فئة "قطاع خاص"، تليها فئة "قطاع عام".
3. كانت فئة الدخل "<2000 ريال" أكثر فئات الدخل مبيتاً بمنى، تليها فئة الدخل "2000-4000".
4. كانت الفئة العمرية "40-50" هي أكثر الفئات العمرية مبيتاً بمنى ، تليها الفئة العمرية "30-40".

ويظهر شكل (6) التمثيل المرئي لتلك العلاقات.



شكل (6): التمثيل المرئي للعلاقة بين الجنسية والوظيفة والدخل وعدد ليالي البقاء في منى.

1. **دراسة العلاقة بين المتغيرات**

من المهم معرفة العلاقة بين المتغيرات ودراستها وذلك بوضع المتغير موضع الدراسة *كمتغير هدف* ودراسة تأثير المتغيرات الأخرى عليه لمعرفة ترتيب تأثير تلك المتغيرات على المتغير الهدف أو ما يسمى ب " Attribute Ranking"، وقد قمنا بدراسة ذلك باستخدام خوارزمية "Gain Ratio Attribute Evaluator [2]".

وتمت دراسة العلاقة بين **إنفاق الحاج** و**عدد مرات الحج** و**عدد** **ليالي المبيت في منى** و**عدد** **أيام البقاء في المدينة** **المنورة** كمتغيرات هدف مع باقي المتغيرات، فكانت المتغيرات المؤثرة حسب قوة تأثيرها مرتبة كما هو مبين في الجداول (1-6) ~ (4-6).

جدول (1-6): المتغيرات المؤثرة في إنفاق الحاج.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| المتغير الرابع | المتغير الثالث | المتغير الثاني | المتغير الأول | المتغير الهدف |
| اللغة(0.15) | الجنسية(0.1782) | المبلغ المدفوع للشركة (0.254) | الوظيفة(0.3609) | الإنفاق |

جدول (2-6): المتغيرات المؤثرة في عدد مرات الحج.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| المتغير الرابع | المتغير الثالث | المتغير الثاني | المتغير الأول | المتغير الهدف |
| الإنفاق(0.1332) | المبلغ المدفوع للشركة(0.1446) | الجنسية(0.1601) | الوظيفة(0.2012) | عدد مرات الحج |

جدول (3-6): المتغيرات المؤثرة في عدد ليالي المبيت في منى.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| المتغير الرابع | المتغير الثالث | المتغير الثاني | المتغير الأول | المتغير الهدف |
| الإنفاق(0.0732) | الوظيفة(0.0783) | الجنسية(0.1466) | اللغة(0.1496) | عدد ليالي المبيت في منى |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| المتغير الرابع | المتغير الثالث | المتغير الثاني | المتغير الأول | المتغير الهدف |
| الوظيفة (0.2826) | اللغة (0.3327) | المبلغ المدفوع للشركة (0.5118) | الجنسية(0.5221) | عدد أيام البقاء في المدينة |

جدول (4-6): المتغيرات المؤثرة في عدد أيام البقاء في المدينة المنورة.

ويمكن تلخيص النتائج السابقة كالتالي:

1. تم اختيار بعض المتغيرات (كمتغير هدف) مع المتغيرات الأخرى لدراسة العلاقة بينهم، فكانت النتائج كالتالي:
	1. **الإنفاق** والجنسية والعمر: كانت أكثر الجنسيات إنفاقاً على الترتيب هي السعودية واليمن ومصر وتركيا، وأكثر الأعمار إنفاقاً هي ’40-50‘و’50-60‘.
	2. **عدد مرات الحج** والجنسية والعمر: فكانت الجنسيات الأكثر تكراراً للحج على الترتيب هي السعودية والهند وباكستان ومصر واليمن.
	3. **عدد أيام البقاء في المدينة المنورة** والجنسية والدخل: كانت مدة البقاء بالمدينة المنورة على الترتيب 8، 7، 5 أيام والجنسيات الأكثر مكوثاً هم تركيا والهند وباكستان والسودان واليمن، وفئات الدخل الأكثر مكوثاً على الترتيب ’2000-4000‘، ’10000<‘، ’4000-6000‘.
	4. **عدد ليالي المبيت في منى** والجنسية والعمر والوظيفة والدخل:
		1. كانت الجنسية التركية هي أكثر الجنسيات مبيتاً لليلة واحدة، والهند لليلتين، وباكستان لثلاث ليالي.
		2. كانت أكثر المهن مبيتاً بمنى هي فئة "قطاع خاص"، تليها فئة "قطاع عام".
		3. كانت فئة الدخل "<2000 ريال" أكثر فئات الدخل مبيتا بمنى ، تليها فئة الدخل "2000-4000".
		4. كانت الفئة العمرية "40-50" هي أكثر الفئات العمرية مبيتاً بمني، تليها الفئة العمرية "30-40".
2. كما تمت دراسة العلاقة بين المتغيرات وتأثير بعضها على بعض، فكانت النتائج كالتالي:
	1. **عدد مرات الحج** (متغير هدف) تأثر بالوظيفة، الجنسية، المبلغ المدفوع للشركة، ثم الإنفاق على الترتيب.
	2. **عدد ليالي المبيت في منى** (متغير هدف) تأثر باللغة، الجنسية، الوظيفة، ثم الإنفاق على الترتيب.
	3. **عدد أيام البقاء في المدينة المنورة** (متغير هدف) تأثر بالجنسية، المبلغ المدفوع للشركة، اللغة، ثم الوظيفة.

**الخلاصة:**

في هذه الدراسة تم استخدام تقنية التنقيب في البيانات الخاصة بالسجل التاريخي لبيانات الحجاج من موسم حج 1432 إلى موسم حج 1438 كنموذج، بهدف إيجاد أهم العلاقات بين المتغيرات لاستخراج المعلومات المهمة وتمثيلها مرئياً، فكانت أكثر الجنسيات إنفاقاً هي السعودية واليمن ومصر، وأكثر الأعمار إنفاقاً هي الفئة العمرية ’40-50‘و’30-40‘ ، ثم ’50-60‘، وأظهرت عملية التحليل أن الجنسيات الأكثر تكراراً للحج هي السعودية ثم الهندية والباكستانية، والجنسيات الأكثر مكوثاً بالمدينة المنورة هي التركية والهندية والباكستانية، وتمثل فئات الدخل الأكثر مكوثاً بالمدينة المنورة هي فئات الدخل من’2000-4000‘، ’10000<‘ وكانت الجنسيات الأكثر مبيتاً في منى هي التركية والهندية والباكستانية.

ثم تطرقت الدراسة إلى دراسة العلاقة بين المتغيرات الهدف والمتغيرات الأخرى لمعرفة قوة تأثيرها على المتغير الهدف، فعدد مرات الحج تأثرت ب (الوظيفة، الجنسية، المبلغ المدفوع للشركة، ثم الإنفاق)، وعدد ليالي المبيت في منى تأثرت بــ (اللغة، الجنسية، الوظيفة، ثم الإنفاق)، وعدد أيام البقاء في المدينة المنورة تأثرت بــ (الجنسية، المبلغ المدفوع للشركة، اللغة، ثم الوظيفة).

ومن خلال هذه الدراسة نوصي بعمل دراسات أكثر استفاضة للاستفادة من النواحي الاقتصادية والسياحية والخدمية والتوعوية لخدمة ضيوف الرحمن، فعلى سبيل المثال لا الحصر يمكن عمل دراسة على نوعية الإنفاق للجنسيات والفئات الأكثر إنفاقاً لتوفير متطلباتهم، بالإضافة الى استحداث دراسة لإيجاد حزم تشجيعية للحجاج الأكثر تكراراً للحج، والقيام بدراسة توعوية لأكثر الجنسيات مبيتاً بمنى.

**التوصيات:**

1. عمل دراسة اقتصادية لنوعية الإنفاق لجنسيات الحجاج الأكثر إنفاقاً وفئاتهم لتوفير متطلباتهم الشرائية والغذائية والسكنية وذلك بالتعاون مع برنامج خدمة ضيوف الرحمن.
2. عمل دراسة توعوية لمعرفة الجنسيات والفئات الأكثر مبيتاً بمنى لتوعيتهم لمنع الافتراش والتكدس بمنى والتي تهم وزارة الحج والعمرة والجهات الأمنية.
3. عمل دراسة لمعرفة الحجاج الأكثر دخلاً وإنفاقاً لتشجيع مجال الانفاق في النواحي السياحية والتجارية لإنعاش الاقتصاد الداخلي للمملكة.

**المراجع:**

* B. Aydın, G. Pataki, H. Wang, A. Ladha, E. Bullitt, J.S. Marron(2018). “Visualizing the Structure of Large Trees”, 2018.
* S. Barak, and M. Modarres, developing an approach to evaluate stocks by forecasting effective features with data mining methods, Expert Systems with Applications, 2015.
* D. Berend, and A. Kontorovich(2015). “A Finite Sample Analysis of the Naive Bayes Classifier”, Journal of Machine Learning Research 16 (2015) 1519-1545.
* ["Data Mining Curriculum"](http://www.kdd.org/curriculum/index.html). [ACM](https://en.wikipedia.org/wiki/Association_for_Computing_Machinery) [SIGKDD](https://en.wikipedia.org/wiki/SIGKDD). 2006-04-30. Retrieved 2014-01-27.
* Clifton, Christopher (2010). ["Encyclopædia Britannica: Definition of Data Mining"](http://www.britannica.com/EBchecked/topic/1056150/data-mining). Retrieved 2010-12-09.
* S. Dinakaran, P. Thangaiah(2015). “Role of Attribute Selection in Classification Algorithms”, International Journal of Scientific & Engineering Research, Volume 4, Issue 6, June-2015.
* Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (2008). "From Data Mining to Knowledge Discovery in Databases" (PDF). Retrieved 17 December 2008.
* M. Giri, S. Soni(2013). “ Effectiveness of Software Development Process Using Programmer Ranker Algorithm in Pair Programming”, INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, 2013.
* [Hastie, Trevor](https://en.wikipedia.org/wiki/Trevor_Hastie); [Tibshirani, Robert](https://en.wikipedia.org/wiki/Robert_Tibshirani); [Friedman, Jerome](https://en.wikipedia.org/wiki/Jerome_H._Friedman) (2009). ["The Elements of Statistical Learning: Data Mining, Inference, and Prediction"](https://web.archive.org/web/20091110212529/http%3A/www-stat.stanford.edu/~tibs/ElemStatLearn/). Archived from [the original](http://www-stat.stanford.edu/~tibs/ElemStatLearn/) on 2009-11-10. Retrieved 2012-08-07.
* Han, Kamber, Pei, Jaiwei, Micheline, Jian (June 9, 2011). [Data Mining: Concepts and Techniques](http://proquest.safaribooksonline.com.proxy.library.carleton.ca/book/databases/data-warehouses/9780123814791) (3rd ed.). Morgan Kaufmann. [ISBN](https://en.wikipedia.org/wiki/International_Standard_Book_Number) [978-0-12-381479-1](https://en.wikipedia.org/wiki/Special%3ABookSources/978-0-12-381479-1).
* R (programming language), The R Project for Statistical Computing, 2018.
* RapidMiner, <https://rapidminer.com/>, 2018.
* Visualization of large tree structures, http://www.randelshofer.ch/treeviz/, 2018.
* Weka (machine learning), <https://www.cs.waikato.ac.nz/~ml/weka/>, 2018.