

المقارنة بين النموذج اللوجستي الثنائي

ونماذج الشبكات العصبية الاصطناعية للتمييز بين دخل الأسرة

إعداد

1/ أ. علي أبشر فضل المولى سليمان

باحث- كلية العلوم - قسم الإحصاء التطبيقي /جامعة السودان للعلوم و التكنولوجيا

ت: 0915024955 – 0115241113 Email: aliabbashar@yahoo.com

2/ د. أحمد محمد عبد الله حمدي

استاذ مشارك / كلية العلوم / جامعة السودان للعلوم والتكنولوجيا

ت : 0117711937 Email: hamdiam@yahoo.com

Abstract

In this study comparison between artificial Neural Networks and Binary Logistic Regression Model has been undertaken, and that is to classify observations to the group that are belonged to, in case when some variables don't follow the normal distribution. That is for preference between the two used methods. We use the proportion of misclassified observations as a criterion to compare between the two methods. The problem lies in how to differentiate and classify the incomes of families in terms of income adequacy and insufficient income under the large increase in the prices of most essential goods which constitute a pressure on the income of the household head and does not cover all the needs of the family, as well as a multiplicity of methods of statistical processing of the data quality and the conditions of use of each of them. The study aimed at knowing the most important variables that affect the sufficiency of household income, and comparison between the Binary Logistic model and the Neural Networks models, in addition to the recognition of the best statistical method between the two methods for processing such data. The study has used the descriptive and inductive analytical method, and that is by analyzing the study data and thereafter deriving the results. The study used the Statistical Package for Social Science (SPSS) version 20 for processing the data. The main results that the study reached that the two used methods are similar regarding significance of both the effect and the importance of the independent variables that considered in the analysis. But the method of artificial neural

networks gained a better classification proportion than the Binary Logistic Regression model. The most important recommendations of the study that making use of the statistical methods in all fields of knowledge, and generalizing the application of both Neural Networks and Logistic model in all fields of knowledge.

المستخلص

في هذه الدراسة تمت المقارنة بين طريقة الشبكات العصبية ونموذج الإنحدار اللوجستي، لتصنيف المشاهدات إلى المجموعة التي تنتمي إليها في حالة كون بعض المتغيرات لا تتبع التوزيع الطبيعي، وذلك للمفاضلة بين الطريقتين المستخدمتين. وقد تم استخدام معيار نسبة المشاهدات المصنفة خطأ (نسبة التصنيف الخاطئ) كمعيار للمقارنة بين الطريقتين. تمثلت مشكلة الدراسة في كيفية تمييز وتصنيف دخول الأسر من حيث الكفاية إلى دخل كافي ودخل غير كافي في ظل الزيادة الكبيرة في أسعار معظم السلع الضرورية مما يشكل ضغط على دخل رب الأسرة وتجعله لا يغطي كل إحتياجات الأسرة، بالإضافة إلى تعدد الطرق الإحصائية لمعالجة البيانات النوعية وشروط استخدام كل منها. هدفت الدراسة إلى التعرف على أهم المتغيرات التي تؤثر على كفاية دخل الأسرة، والمقارنة بين أسلوب النموذج اللوجستي الثنائي ونماذج الشبكات العصبية، هذا بالإضافة إلى التعرف على الأسلوب الإحصائي الأفضل بين الأسلوبين لمعالجة مثل هذه البيانات. استخدمت الدراسة الأسلوب الوصفي والاستقرائي التحليلي وذلك بتحليل بيانات الدراسة ومن ثم استنباط النتائج. استخدمت الدراسة برنامج الحزمة الإحصائية للعلوم الإجتماعية الإصدار 20 (SPSS version 20) لمعالجة البيانات. من أهم نتائج الدراسة أن الطريقتين المستخدمتين متطابقتين من حيث معنوية كل من تأثير وأهمية المتغيرات المستقلة الداخلة في التحليل إلا أن طريقة الشبكات العصبية الإصطناعية قد أعطت نسبة تصنيف أفضل من نموذج الإنحدار اللوجستي الثنائي. من أهم توصيات الدراسة: الاستفادة من الأساليب الإحصائية في جميع مجالات المعرفة، وتعميم تطبيق أسلوب الشبكات العصبية والنموذج اللوجستي في جميع مجالات المعرفة.

المقدمة

لعملية الفصل والتمييز بين المشاهدات يستخدم أسلوب الدالة التمييزية بصورة شائعة لهذا الغرض أو النموذج اللوجستي إذا كانت المتغيرات المستقلة خليط ما بين المتقطعة و المتصلة وفي الآونة الأخيرة برز استخدام الشبكات العصبية الإصطناعية Artificial Neural networks لأغراض التمييز والفصل بين المجموعات، كما يستخدم أسلوب الشبكات العصبية الإصطناعية في

جوانب تطبيقية أخرى مثل التنبؤ والسيطرة النوعية ومحاكاة عمل الرجل الآلي وتحليل الصور، وذلك بسبب المرونة التي يتمتع بها أسلوب الشبكات العصبية الاصطناعية وقدرته على التعامل مع الدوال غير الخطية لأنه لا يعتمد على نوع التوزيع التي تتبعه متغيرات الظاهرة المدروسة، فأسلوب الشبكات العصبية هو أسلوب تكراري يهدف إلى جعل الخطأ أقل ما يمكن من خلال تعليم Learning أو تدريب Training الشبكة العصبية بهدف الوصول إلى الأوزان Weights التي تؤدي إلى تقليل الخطأ ، أما النموذج اللوجستي فهو من النماذج الأسية التي يمكن تحويلها إلى خطية باستخدام دالة أو رابط معين، يستخدم الإنحدار اللوجستي لمعالجة النماذج ذات المتغيرات التابعة النوعية ، مثل عملية التصنيف إلى مجموعتين (ثنائي) أو أكثر ، ففي هذه الدراسة ستتم المقارنة بين أسلوبين النموذج اللوجستي الثنائي ونماذج الشبكات العصبية الاصطناعية .

لقد كانت موجة الإهتمام بأسلوب الشبكات العصبية بعد إعطاء الباحثان McCulloch و Pitts عام 1943 مقدمة في الخلايا العصبية Neurons كخلايا بيولوجية ، وكيفية محاكاة عمل هذه الخلايا لإجراء عمليات حسابية، ولكن لم يتم استخدام أسلوب الشبكات العصبية بكثرة في السنوات الأولى للإكتشاف بسبب صعوبة تطبيقها وغموضها بالنسبة للباحثين ، وقد برزت موجة الاستخدام للشبكات العصبية بعد اقتراح أسلوب تدريب جديد يعتمد علي الإنتشار العكسي للخطأ ، من قبل مجموعة من الباحثين (Hinton وآخرون) عام 1986م والذي أخذ مجالاً واسعاً في التطبيقات العملية حتى الآن.

لقد تم استخدام الشبكات العصبية لأغراض التمييز في العديد من البحوث على الصعيد الدولي ففي العام 1990م وجد الباحثان Surkan و Singleton إن أسلوب الشبكات العصبية تعطي نتائج أفضل من التحليل التمييزي التقليدي في التنبؤ المستقبلي لتقويم الخطورة في السنوات المالية(1) ، وكذلك في العام 1991م قام الباحثان Wiss و Kulikowski (1) بتقديم طرقاً للتمييز مكونة من طرائق تقليدية وطرق الشبكات العصبية ولأنواع مختلفة منها ، وفي العام 2008م استخدم الباحث طاهر ريسان دخيل أسلوب الدالة التمييزية الخطية و الشبكات العصبية وتوصل إلى أن أسلوب الشبكات العصبية تعطي نتائج أفضل من الدالة التمييزية الخطية (1)، وفي العام 2008م أيضاً قام الباحثان محمد عبد الرحيم حمدي وكرم حاتم ذنون باستخدام أسلوب الشبكات العصبية الاصطناعية في تمييز المصابين بمرض السل الرئوي وتوصلا إلى أن استخدام الشبكات العصبية المتمثل في شبكة Hamming & Maxent تعطي نتائج تمييز جديدة.

مشكلة الدراسة: تكمن مشكلة الدراسة في كيفية تمييز وتصنيف دخول الأسر من حيث الكفاية إلى دخل كافي ودخل غير كافي في ظل الزيادة الكبيرة في أسعار معظم السلع الضرورية مما يشكل ضغط على دخل رب الأسرة وتجعله لا يغطي كل إحتياجات الأسرة ، بالإضافة إلى تعدد الطرق الإحصائية لمعالجة البيانات النوعية وشروط استخدام كل منها.

أهداف الدراسة: تهدف هذه الدراسة إلى الأتي: التعرف على أهم المتغيرات التي تؤثر على كفاية دخل الأسرة. المقارنة بين اسلوبى الشبكات العصبية الإصطناعية و الإنحدار اللوجستي الثنائي في التمييز بين دخول مجموعة من الأسرة إلى دخل كافي ودخل غير كافي . التعرف على الأسلوب الإحصائي الأفضل من بين الأسلوبين المذكورين لتصنيف بيانات الدراسة ومعالجة مثل هذه البيانات.

منهجية الدراسة: بهدف تحقيق الأهداف المنشودة من هذه الدراسة، فقد اعتمد الدارس على المنهج التحليلي الاستنباطي وذلك من خلال تحليل البيانات باستخدام أسلوب النموذج اللوجستي الثنائي وأسلوب الشبكات العصبية الاصطناعية ومقارنة النتائج، من ثم استنباط بع المؤشرات والعلاقات، وكذلك استخدام المنهج الوصفي من خلال توصيف نموذج الإنحدار اللوجستي الثنائي و نماذج الشبكات العصبية الإصطناعية.

مجتمع وعينة الدراسة: يتمثل مجتمع الدراسة في جميع الأسر بولاية الخرطوم، أما عينة الدراسة فتكونت من 545 رب أسرة أو ربة أسرة، تم أخذها عن طريق العينة العشوائية متعددة المراحل.

الأدوات الإحصائية المستخدمة: للمقارنة بين نتائج النموذج اللوجستي الثنائي ونماذج الشبكات العصبية سيتم الاعتماد على نسبة التصنيف الخاطئ لكل نموذج على حده ، وذلك باستخدام برامج الحزم الإحصائية للعلوم الاجتماعية SPSS إصدار 20 باستخدام دالة النموذج اللوجستي الثنائي Binary Logistic ، ودالة نماذج الشبكات العصبية Neural Network من قائمة Analyze باستخدام شبكة البرسيبترون متعددة الطبقات Multilayer Perceptron.

نموذج الانحدار اللوجستي الثنائي البسيط

Simple Binary Logistic Regression Model:-

1- مفهوم نموذج الإنحدار اللوجستي الثنائي

إذا كان (Y) متغير تابع يأخذ القيمة (واحد) إذا تحقق حدث معين والقيمة (صفر) إذا لم يتحقق

ذلك الحدث أي عندما يكون للمتغير التابع ذو قيمتين فقط (0,1)، وأن (X) متغير كمي أو نوعي، يسمى هذا النموذج بنموذج الانحدار اللوجستي الثنائي Binary Logistic Regression Model، فإذا كان هنالك متغير مستقل واحد يعرف النموذج بالنموذج اللوجستي الثنائي البسيط، أما إذا كان لدينا أكثر من متغير مستقل فإن النموذج يعرف بالنموذج اللوجستي الثنائي المتعدد.

بما أن قيم (E(Y)) تتحصر بين (الصفر والواحد) وبالتالي تتراكم بين هاتين القيمتين كما يعتمد شكل الدالة التزايدية والتناقصية على إشارة (β_j) والمتغير التابع في النموذج المقدر هو متغير برنولي يأخذ إحدى القيمتين: القيمة (1) باحتمال P والقيمة (0) باحتمال 1-p.

دالة هذا المنحنى هي الدالة اللوجستية، لهذا فعندما يكون (Y) متغير ثنائي فإن القيمة المتوقعة للمتغير التابع بمعلومية المتغير المستقل يعبر عنها كالتالي:

$$E(y/x) = \frac{e^{(\beta_0 + \beta_1 X)}}{1 + e^{(\beta_0 + \beta_1 X)}} = P(x) \dots \dots \dots (1)$$

ويعمل رياضي بسيط على المعادلة السابقة نجد أن:

$$\frac{P(x)}{1-P(x)} = e^{(\beta_0 + \beta_1 X_i)} \dots \dots \dots (2)$$

وتسمى المعادلة السابقة بالدالة اللوجستية الثنائية أو النموذج اللوجستي الثنائي البسيط

Simple Binary Logistic Regression Model

ومن خصائص الدالة اللوجستية أنه يمكن تحويلها لدالة خطية وذلك بأخذ اللوغاريتم للطرفين: كالتالي :

$$\ln \left\{ \frac{\hat{P}(x)}{[1-\hat{P}(x)]} \right\} = \beta_0 + \beta_1 X_i \dots \dots \dots (3)$$

(P(x) هي احتمال وجود الظاهرة، بينما (1 - P(x)) هي احتمال عدم وجود الظاهرة .

وتسمى g(x) اللوجت Logit وترجع أهمية النموذج اللوجستي لكون المقدار (e^{β₁}) يعطي ما يسمى بنسبة الخطر Odd Ratio.

2- النموذج اللوجستي المتعدد:

إذا كان لدينا متغير تابع (Y) ثنائي يأخذ القيم (صفر وواحد) و P من المتغيرات المستقلة.

$$X = (x_1, x_2, x_3, \dots, x_p) \dots \dots \dots (4)$$

يمكن تعميم النموذج في (2-1-3) ليأخذ الشكل :

$$E(y/x) = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p)}} \dots (5)$$

وفي هذه الحالة يسمى بالنموذج اللوجستي المتعدد كما في المعادلة (3-6) وتأخذ دالة اللوجت الشكل التالي:

$$g(X_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p \dots (6)$$

إذا كان أحد المتغيرات المستقلة نوعي أو اسمي وله K من المستويات عندئذ سيكون له K - 1 من الحالات نرسم لها بالرمز D_{ji} ولدنا $i = 1, 2, 3, \dots, K-1$ حينها تصبح المعادلة :

$$g(X_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \sum_{i=1}^{K_j-1} \beta_{ji} D_{ji} + \beta_p X_p \dots (7)$$

ومن النموذج (7) نجد أن قيمة واحدة فقط من D_{ji} تأخذ القيمة (واحد) وذلك عندما يتصف المتغير المستقل بصفة معينة، بينما باقي القيم تأخذ القيمة (صفر) دلالة على عدم وجود الصفة للمتغير المستقل.

3- تقدير معاملات الإنحدار اللوجستي المتعدد :

الطريقة التي تستخدم في تقدير معاملات النموذج اللوجستي هي طريقة الترجيح الأعظم Maximum Likelihood وهي من الطرق الشائعة لتقدير معاملات النماذج الغير خطية، وتعتبر أفضل من طريقة المربعات الصغرى العادية المستخدمة لتقدير معاملات النماذج الخطية لأنها لا تفترض قيود كثيرة مثل الخطية وثبات التباين، وهي تقيس الاحتمالات المشاهدة لعدد n من المتغيرات المستقلة في العينة، أي أن مضاريب تلك الاحتمالات يمثل دالة الإمكان الأعظم.

تعتبر طريقة الترجيح الأعظم من أكثر طرق التقدير استعمالاً في تقدير معاملات النماذج المجهولة ، فإذا كانت لدينا عينة X عشوائية مختارة من توزيع احتمالي $f(x_i, \beta)$ فإن التوزيع المشترك لهذه العينة هو :

$$F(\underline{x}, \beta) = F(x_1, x_2, x_3, \dots, x_n) = \prod_{i=1}^n f(x_i, \beta)$$

ويتلخص مبدأ دالة الترجيح الأعظم في إيجاد المقدّر $\hat{\beta}$ الذي يجعل دالة الترجيح أكبر ما يمكن⁽¹⁾ ، أي أن مقدّر الإمكان للمعلمة β هو القيمة $\hat{\beta}$ التي يحقق العلاقة:

$$F(\underline{x}, \hat{\beta}) \geq F(\underline{x}, \beta) \dots\dots\dots (8)$$

ويُفسر ذلك التقدير المتقطعة على أنه قيمة (β) التي تجعل احتمال سحب العينة المشاهدة أكبر ما يمكن ، أما بالنسبة للتوزيعات المتصلة فإن قيمة (β) التي تجعل الإحتمال أكبر ما يمكن للحصول على قيم للعينة قريبة جداً من القيم التي حصلنا عليها ، فإن الدالة التي تحقق المعادلة رقم (8) أعلاه تسمى بدالة الترجيح الأعظم Maximum Likelihood

4- إختبار معنوية معلمات النموذج:

عند بناء النموذج نجد أن من المفيد الإجابة على السؤال هل يمكن حذف أي من المتغيرات في نموذج الإندثار؟ وللإجابة نقوم بإجراء إختبار ما إذا كانت معلمات الإندثار (β_j) تساوي صفر أم لا ؟ وإختبار ذلك نحسب أولاً ما يسمى بالإندثار (Deviance) ويرمز له بالرمز (D) والإندثار لوغريثم النسبية بين دالة الإمكان ودالة النموذج المشبع (Saturated Model) مضروبة في 2، ويقصد بالنموذج المشبع النموذج الذي يحوي عدداً من المعلمات مساوياً لعدد المشاهدات وهو يوفق البيانات تماماً وتكون قيمته واحد صحيح. يقوم الفرق (deviance(D) على أساس الإختلاف بين قيمتي لوغريثم الترجيح مع الملاحظ إنه لا يمكن أن تكون قيمة لوغريثم الترجيح للنموذج المقدّر لديه معلمات أقل.

يمكن إسترجاع نموذج الفرق (Deviance model) كشكل يوضح جودة التوفيق (Goodness of Fit) وكلما كان الفرق كبير فهذا يدل على عدم المطابقة وعدم الجودة.

أي أن الإندثار يأخذ الشكل :

$$D = -2 \log \left[\frac{\text{LikeLihood Of The Fitting model}}{\text{LikeLihood Of The Saturated model}} \right] \dots\dots\dots (9)$$

حيث :

Log : هو اللوغريثم الطبيعي

نجد أن قيمة D شبيهة بقيمة SSE في الإندثار الخطي حيث نجد أن

(1) جلال مصطفى الصياد (1993م) ، الاستدلال الإحصائي ، ص 197.

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \dots\dots\dots(10)$$

ولمعرفة أهمية وجود المتغيرات المستقلة في النموذج نقارن قيمة D للنموذج متضمن المتغيرات مع قيمة D عندما يكون النموذج غير متضمن لتلك المتغيرات وذلك كما يلي:

$$G = D(\text{model without variable}) - D(\text{model with variable})$$

نجد أن G تلعب نفس قيمة F في الإنحدار الخطي وبما أن دالة الإمكان للنموذج المشبع تساوي (واحد) بالتالي الصيغة كالآتي:

$$G = -2 \log_e \left[\frac{\text{LikeLihood Without the variable}}{\text{LikeLihood with the variable}} \right] \dots\dots\dots(11)$$

$$= -2 \log_e \text{ LikeLihood without the variable} + 2 \log_e \text{ LikeLihood with the variable}$$

وتعطي جميع البرمجيات قيم لوغريثم دالة الإمكان (مضروبة في 2) لأي نموذج.

في حالة صحة إختبار فرضية العدم بأن معاملات المتغيرات تساوي صفر (أي أن المتغيرات غير ذات أهمية) نجد أن G تتبع توزيع X^2 بدرجات حرية (P) (حيث P عدد المتغيرات). لهذا نقارن قيمة G المشاهدة بالقيمة الحرجة من جدول X^2 بدرجات حرية (P) فإذا وجدنا أن القيمة المشاهدة أكبر من الحرجة نرفض فرض العدم.

5- إختبار Wald :

يستخدم هذا الإختبار لمعرفة معنوية معلمة معينة في النموذج حيث أنه يختبر فرض العدم التالي:

$$H_0 : \beta_j = 0$$

ضد الفرض البديل:

$$H_1 : \beta_j \neq 0$$

ويستند الإختبار على الإحصائية:

$$Z = \frac{\hat{\beta}_j}{\delta E(\hat{\beta}_j)} \dots\dots\dots(12)$$

حيث $s.E(\hat{\beta}_j)$ مقدر الخطأ المعياري للمعاملات $\hat{\beta}_j$ والتي تتبع (في حالة صحة H_0)

التوزيع الطبيعي، لهذا نقارن قيمتها المشاهدة بالقيم الحرجة ولا نرفض H_0 إذا كانت تقع بينهما.

6- إختبار Hosmer – Lemshow

يستخدم الإختبار لمعرفة مدى مطابقة البيانات للنموذج المقترح، وذلك بإختبار الفرضية الآتية:

H_0 : النموذج مطابق لبيانات النموذج اللوجستي المتعدد

H_1 : النموذج غير مطابق

حيث تقارن قيمة هذا الإختبار مع القيمة الجدولية لمربع كاي X^2 فإذا كانت القيمة المحسوبة أقل أو تساوى الجدولية هذا يعنى أننا نقبل فرض العدم أي أن النموذج مطابق للبيانات، أما إذا كانت القيمة المحسوبة أكبر فهذا يعنى رفض العدم وقبول الفرض البديل والذي يفرض عدم مطابقة البيانات للنموذج.

الشبكات العصبية الاصطناعية (ANN): Artificial Neural Networks

1- مفهوم الشبكات العصبية الاصطناعية:

إن الشبكات العصبية الاصطناعية ANN هي عبارة عن صيغ رياضية تعتمد على نماذج رياضية تحاكي عمل الدماغ البشري في حل المشكلات وإجراء الحسابات ، ويطلق على الشبكات العصبية أيضاً عدة مسميات مثل الأنظمة المترابطة Connection Systems و الأنظمة التكيفية Adaptive Systems أو أنظمة التوزيع المتوزي Parallel Distribution Systems ، وإن السر من هذه التسميات المتعددة هي بسبب أن الشبكات العصبية الاصطناعية مكونة من سلسلة من عناصر المعالجة وهي الخلايا العصبية Neurons الاصطناعية التي تحاكي الشبكات العصبية البيولوجية في الإنسان، حيث تكون تلك العناصر مترابطة فيما بينها وفق خطوط متوازية، بحيث أن جميع هذه العناصر تتغير وتُحدث على وفق قواعد وطرائق تكيفية معينة و التي بدورها تعتمد على المعلومات المدخلة في الشبكة العصبية الاصطناعية ANN.

إن من أهم أهداف الشبكات العصبية الاصطناعية ANN هي فهم كيفية تكوين و تشكيل المميزات الدالية و الخصائص الحسابية للدماغ عندما ينجز بعض العمليات مثل الإدراك الحسي و التمييز بين الاشياء و التعليم ومحاولة الإستفادة من هذه الخصائص لغرض تحويلها بإتجاه رياضي

وحسابي قادر على تحليل البيانات و المعلومات ومن ثم التنبؤ ، وفي الوقت الحاضر فإن هناك العديد من المحاولات التي تبذل في سبيل تقدير تطبيقات الشبكات العصبية و الإصطناعية ANN واستخدامها في العديد من المجالات مثل تحليل السلاسل الزمنية و التحليل التمييزي وغيرها.

2- هيكل الشبكة العصبية الإصطناعية Artificial Neural Network Structure

نقصد بهيكل الشبكة العصبية مكونات الشبكة العصبية الإصطناعية ، تتكون الشبكة العصبية الإصطناعية من مجموعة من وحدات المعالجة و التي تسمى بالخلايا العصبية أو عصبونات والتي تشابه العصبونات البيولوجية الموجودة في مخ الإنسان ، وهذه الوحدات متصلة فيما بينها في شكل تراكيبات تسمى بالشبكة العصبية.

حيث تعمل الروابط على تمرير الإشارات من عصبون إلى آخر ويكون لكل رابط وزن عددي يصاحبه ، كما يستقبل كل عصبون عدد من المدخلات عن طريق تشابكاته وينتج إشارة مخرجات واحدة فقط ، حيث تنتقل إشارة المخرجات عن طريق الرابط الخارجي للعصبون ، وينقسم الرابط الخارجي بدوره إلى عدد من الفروع التي تنقل نفس الإشارة. فمجموعة هذه المكونات تمثل البنية العامة أو الشكل العام للشبكات العصبية بحيث أن أي شبكة عصبية تتكون من عدة طبقات - طبقتين على الأقل - زائداً الوصلات البينية أي تتكون الشبكة من العناصر التالية:

طبقة المدخلات ، طبقة المخرجات ، الطبقة الخفية ، الوصلات البينية.

3- البنية المعمارية للشبكة العصبية Architecture of ANN:

بنية الشبكة أو معمارية الشبكة العصبية هي الطريقة التي تربط بها وحدات المعالجة مع بعضها البعض داخل كل طبقة أو بين الطبقات المختلفة المكونة للشبكة، حيث ترتبط هذه الوحدات بطرق مختلفة ووفقاً لكيفية هذا الترابط وعدد الطبقات المكونة للشبكة تظهر لنا البنية أو المعمارية العامة للشبكة العصبية. ويمكن تصنيف تراكيب الشبكات العصبية وفقاً لعدد الطبقات إلى:

أ. شبكات وحيدة الطبقة Single Layer Network

ب. شبكات متعددة الطبقات Multilayer Network

أ/ شبكات وحيدة الطبقة:

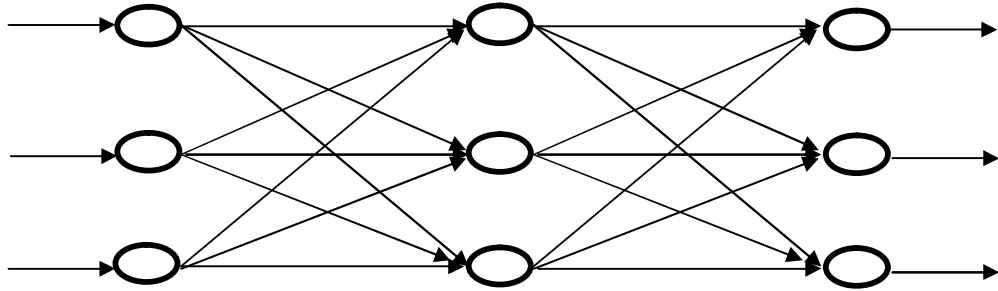
وهي من أبسط أنواع تراكيب الشبكات العصبية وتتألف عادة من طبقة واحدة من عناصر المعالجة حيث ترتبط ربطاً مباشراً مدخلات الشبكة مع مخرجاتها فيتم إجراء كل الحسابات في طبقة

المخرجات وعادة يتم حساب عدد الطبقات في الشبكة بدون حساب طبقة المدخلات لأنها لا تقوم بإنجاز أي حسابات.

ب/ شبكات متعددة الطبقات:

تحتوي تراكيب هذه الشبكة على أكثر من طبقة من عناصر المعالجة التي تربط بينها الوصلات البينية (الأوزان)، حيث تتكون الشبكة متعددة الطبقات على الأقل من طبقتين طبقة لاستقبال المدخلات وهي غير محسوبة أو طبقة المخرجات وبين طبقة المدخلات والمخرجات توجد الطبقة الخفية، ويمكن أن تحتوي الشبكة على أكثر من طبقة خفية، ويتوقف ذلك على نوع التطبيق المستخدم فيه الشبكة.

شكل بياني رقم (1) معمارية الشبكة العصبية متعددة الطبقات



ويمكن تصنيف الشبكات العصبية وفقاً لطبيعة انتشار البيانات عبر طبقات وحداتها وترابطها الي الانواع التالية:

1. شبكات التغذية المتقدمة Feed Forward.
2. شبكات التغذية الراجعة Feed Back.
3. شبكات الترابط الذاتي Auto Associative.
4. شبكات ذاتية التنظيم Self Organizing.

4- استخدام الشبكات العصبية للتصنيف:

يمكن استخدام الشبكات العصبية الإصطناعية ANN كأساس للتصنيف وذلك من خلال تحديد دالة تنشيط Activation Function خاصة بهذا الغرض (التصنيف) ودوال التنشيط هي عبارة عن دوال لمعالجة البيانات الداخلة إلى الوحدات أو الخلايا العصبية ، وهناك العديد من دوال التنشيط التي قُدمت من قبل الباحثين في هذا المجال و التي تختلف تبعاً لإختلاف المخرج (الناتج) الذي يتم

الحصول عليه إضافة إلى أهداف الشبكة العصبية التي يراد تحقيقها ، وهنا سوف يتم استخدام دالة الخطوة Step Function وتسمى أيضاً بدالة العتبة threshold function ، وذلك لأنها تناسب الإستخدامات التصنيفية و التمييزية وذلك لأنها تعطي نتيجتين فقط للمخرج (الناتج) أما صفر أو واحد ، كما في الصيغة التالية:

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases} \dots\dots\dots(13)$$

حيث تستخدم هذه الدالة في وحدات طبقة المخرجات ، بينما تستخدم الدالة اللوجستية Sigmoid في خلايا الطبقة المخفية و التي تعطى سيققتها بالشكل التالي:

$$f(x) = \frac{1}{1+e^{-s}} \dots\dots\dots (14)$$

حيث s تمثل المجموع الموزون للمدخلات مضافاً إليه حد التحيز (Bias) الذي نرسم له بالرمز θ ، أي أن المجموع الموزون يعطى بالصيغة التالية:

$$S = \sum_{i=1}^n w_i x_i + \theta , \dots\dots\dots (15)$$

حيث w_i : تمثل الأوزان. تُحسب قيمة s في وحدة الطبقة المخفية اعتماداً على الدالة اللوجستية، بينما في طبقة المخرج تُحسب اعتماداً على دالة الخطوة ، وبذلك فإن مخرج الشبكة العصبية هو 0 أو 1 ، وبالاعتماد على المدخل فإذا كان المخرج (الناتج) 0 هذا يعني إن المشاهدة تتبع للصنف أو المجموعة الأولى ، أما إذا كان المخرج 1 فإن المشاهدة تتبع للصنف أو المجموعة الثانية.

5- التدريب أو التعليم Training or learning في الشبكات العصبية:

إن تقنية التدريب أو التعليم في الشبكات العصبية قد تم إستيحاء فكرتها من خلال قدرة الإنسان على التعلم و التدرب من تجاربه التي يمر بها وبهذا يمكن أن يُعَلِّم نفسه بنفسه إلى أن يصل إلى حالة الاستقرار ، وبنفس الطريقة فإن تعليم الشبكة العصبية و تدريبها بعطيها القدرة على التعلم حتى الوصول إلى المخرجات المرغوب فيها بأقل خطأ ممكن ، وذلك من خلال إعطاء الشبكة العصبية مجموعة من الأمثلة والتي تسمى بعينة التدريب والتي تُحدَّث الأوزان بين الوحدات ، ومن طرق التعلم للشبكة العصبية طريقة تعليم الإدراك الحسي (Perception) وهذه الطريقة تُعَدِّل وتُحدَّث الأوزان بأسلوب تكراري وذلك من خلال إضافة حد التعديل لقيمة الأوزان القديمة. وتنقسم طرق تعلم الشبكة العصبية الي قسمين رئيسيين هما:

أ/ التعليم بواسطة معلم Supervised Learning.

ب/ التعليم بدون معلم (التعلم الذاتي) .Unsupervised Learning

الإطار التطبيقي للدراسة:

متغيرات الدراسة: تشتمل هذه الدراسة على استخدام ثلاثة متغيرات مدخلة (مستقلة) ضمن مجموعتين لكفاية دخل الأسرة وهي عدد أفراد الأسرة (x_1)، طبيعة ملكية السكن (x_2) (مستأجر أم غير مستأجر) و وجود طلبة جامعيين بالأسرة (x_3) (يوجد / لا يوجد) في حين يمثل المتغير التابع مدى كفاية دخل الأسرة لتغطية نفقات المعيشة.

وصف متغيرات الدراسة: العدد الكلي لعينة الدراسة 545 فرد. يوجد بهذه الأسر نسبة 59.4% طالب. 63% مستأجرون، 59.6% يرون أن دخلهم غير كاف. يتم وصف الدخل الكافي بالرقم 1 وغير الكافي بالصففر.

أولاً: التطبيق العملي باستخدام النموذج اللوجستي الثنائي:

الجدول رقم (1) يبين عدد الدورات التكرارية لمشتقة دالة الإمكان الأعظم

Iteration History ^{a,b,c,d}

Iteration		-2 Log likelihood	Coefficients			
			Constant	الأسرة	س1	س6
Step 1	1	705.172	.572	-.103	.316	-.608
	2	704.837	.675	-.114	.333	-.678
	3	704.837	.677	-.114	.333	-.680
	4	704.837	.677	-.114	.333	-.680
a. Method: Enter						
b. Constant is included in the model.						
c. Initial -2 Log Likelihood: 735.174						
d. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.						

يوضح الجدول رقم (7) عدد الدورات التكرارية لمشتقات دالة الإمكان الأعظم للحصول على أقل قيمة لسالب ضعف لوغاريتم دالة الإمكان الأعظم للحصول على التقدير الأمثل لمعاملات

النموذج، وقد حصلنا في الدورة الرابعة لمشتق سالب ضعف دالة الإمكان الأعظم على أقل قيمة لها وهي مساوية (704.174).

توقفنا عند هذه الدورة لأن التغير في ضعف دالة الإمكان أصبح ضعيفاً جداً (أقل من 0.001) وفي حقيقة الأمر فإن التغير في المعاملات المقدرة (β_1, β_2) , β_3 , أصبح ضعيفاً بعد الدورة الثانية، كما يتضح من الجدول السابق أن مقدرات المعلمات في الدورات (2,3,4) متشابهة مع فروقات بسيطة، لذلك تم التوقف عند الدورة الرابعة باعتبار أن معلماتها تمثل أفضل نتيجة يمكن الحصول عليها لأن سالب ضعف دالة الإمكان الأعظم في نهايتها الصغرى.

جدول رقم (2) يوضح النموذج الأمثل المقدر Variables in the Equation

		B	S.E	Wald	df	Sig	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	حجم الأسرة	-.114	.034	11.19	1	.001	.892	.834	.954
	وجود طلبة جامعيين بالأسرة	.333	.185	3.24	1	.072	1.396	.971	2.006
	طبيعة ملكية السكن	-.680	.195	912.0	1	.001	.507	.346	.743
	Constant	.677	.472	2.06	1	.151	1.969		
Variable(s) entered on step: حجم الأسرة , وجود طلبة جامعيين بالأسرة , طبيعة ملكية السكن.									

يوضح الجدول رقم (2) معاملات النموذج الأمثل الذي حصلنا عليه عند الدورة الرابعة ، حيث العمود الأول (B) يمثل معاملات النموذج المقدر بوحدات $\text{Log} \left(\frac{p}{1-p} \right)$ وتصبح معادلة النموذج المقدر كما يلي:

$$\text{Log} \left(\frac{\hat{p}}{1-\hat{p}} \right) = 0.677 - 0.114X_1 + 0.333X_2 - 0.680X_3$$

حيث :

\hat{p} : تمثل متوسط المتغير التابع عند توفر الظاهرة أي احتمال الحصول على دخل كافي للمفردات الجديدة، وتمثل هذه التقديرات العلاقة بين المتغير التابع والمتغيرات المستقلة.

X_1 : متغير حجم الأسرة.

X_2 : متغير وجود طلبة جامعيين بالأسرة.

X_3 : متغير طبيعة ملكية السكن.

حيث يتضح من الجدول رقم (8) السابق أن إشارة معامل تأثير متغير حجم الأسرة سالب وهذا يعني أن كلما زاد حجم الأسرة تقل قيمة المتغير التابع وتقترب من القيمة (0) أي عدم كفاية الدخل ومن خلال قيمة الإنحدار يتضح أنه كلما زاد حجم الأسرة بمعدل فرد واحد أدى ذلك إلى نقصان احتمال كفاية الدخل بمعدل (0.114)، أما معامل إنحدار متغير وجود طلبة جامعيين بالأسرة فينتضح أن الإشارة موجبة وهذا يعني أنه كلما زاد متغير وجود طلبة جامعيين (يوجد = 1 ، لا يوجد = 2) أي اقترب من القيمة (2) زاد احتمال أن يأخذ المتغير التابع القيمة واحد ($p(y) = 1$) أي كفاية الدخل ومن خلال القيمة يتضح أنه كلما اقترب متغير وجود طلبة من القيمة (2) زاد احتمال كفاية الدخل بمعدل (0.333)، أما معامل انحدار طبيعة ملكية السكن (مستأجر = 1، غير مستأجر = 2) فتظهر الإشارة سالبة وهذا يعني أن كلما اقترب متغير ملكية السكن من القيمة (2) غير مستأجر يتجه المتغير التابع لأخذ القيمة (0) عدم كفاية الدخل والعكس صحيح، ومن خلال القيمة يتضح أنه كلما زاد متغير طبيعة ملكية السكن واقترب من القيمة (2) أدى ذلك إلى إنخفاض احتمال كفاية الدخل بمعدل (0.68)، مما سبق نستنتج أن كل إشارات معاملات الإنحدار منطقية وهي تعكس علاقة المتغيرات المستقلة بالمتغير التابع وذلك حسب طريقة الترميز الموضحة.

أما العمود الثاني (S.E) يمثل الخطأ المعياري للمعاملات (المقدرات)، والعمود الثالث إحصاءة Wald لاختبار معنوية المعلمات المقدره فينتضح من الجدول السابق أن المعاملات الخاصة بمتغيري حجم الأسرة وطبيعة ملكية السكن معنويين وذلك من خلال القيم الاحتمالية (Sig) بالعمود الخامس أي لهما تأثير معنوي على كفاية دخل رب الأسرة، أما متغير وجود طلبة جامعيين بالمنزل ليس له تأثير معنوي على كفاية دخل رب الأسرة إذ أن القيمة الاحتمالية الخاصة بالمعلمة (0.072) أكبر من مستوى المعنوية (0.05)، العمود الرابع يمثل درجات الحرية، أما العمود الخامس فيمثل القيم الاحتمالية (Sig) التي على أساسها يتم تحديد معنوية معامل الانحدار، أما العمود السادس ($Exp(p)$) يوضح قيمة الدالة الأسية لمعامل الانحدار وهو يعبر عن المضاعف

الذي تتغير به نسبة الترجيح (قيمة المتغير التابع) أي التغير من احتمال (Y=1) إلى احتمال (Y=0) فيتم حسابها من الصيغة التالية:

$$\text{Exp}(p) = e^{\beta_i}$$

مثلاً القيمة الأولى نحصل عليها كالتالي:

$$\text{Exp}(p) = e^{(-0.114)} = 0.892$$

أما العمود الأخير فيمثل حدود الثقة لقيمة الدالة الأسية التي تم حسابها سابقاً.

الجدول رقم (3) نتائج Chi-square لاختبار معنوية النموذج

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Model	30.337	3	.000

يوضح الجدول رقم (3) السابق نتائج اختبار كاي تربيع (Chi-square) لاختبار معنوية النموذج أي لاختبار فرضية العدم التالية:

النموذج غير معنوي: H_0

فيتضح من الجدول السابق معنوية الاختبار وذلك من خلال القيمة الاحتمالية (Sig) أي وجود فروق ذات دلالة إحصائية بين فرضية العدم والنتائج المشاهدة وبالتالي رفض فرضية العدم (H_0) وقبول الفرض البديل أي أن النموذج بالكامل معنوي ويمثل البيانات تمثيلاً جيداً .

الجدول رقم (4) القيم المشاهدة والمتوقعة لاختبار هوسمر & لمشو

Contingency Table for Hosmer and Lemeshow Test						
		هل تعتقد أن دخلك كافي لتغطية نفقات معيشة الأسرة = غير كافي		هل تعتقد أن دخلك كافي لتغطية نفقات معيشة الأسرة = كافي		Total
		Observed	Expected	Observed	Expected	
Step 1	1	43	45.776	15	12.224	58
	2	35	36.975	16	14.025	51
	3	36	38.666	21	18.334	57
	4	37	34.365	17	19.635	54

5	33	31.257	18	19.743	51
6	43	33.850	15	24.150	58
7	31	26.973	18	22.027	49
8	27	28.883	29	27.117	56
9	22	26.055	34	29.945	56
10	18	22.200	37	32.800	55

الجدول رقم (4) السابق يوضح نتائج اختبار هوسمر & لمشو، وهو اختبار لا معلمي يستخدم أيضاً للتأكد من جودة توفيق النموذج ويعتمد على حساب قيمة إحصاءة (Chi-square) للفرق بين القيم المشاهدة والقيم المتوقعة، وقد اقترح Hosmer and Lemeshow هذا الاختبار للكشف عن انحرافات النموذج اللوجستي أي الفرق المعنوي بين القيم المشاهدة والقيم المتوقعة، وتتكون إحصاءة هذا الاختبار من جزئين: جزء مشاهد (Observed) ويمثل بيانات العينة والجزء الآخر متوقع (Expected) محسوب من النموذج المقدر، وذلك بتجزئة قيم مشاهدات العينة وغالباً تتم تجزئة البيانات إلى 10 أجزاء، كما بالجدول السابق وتستخدم احصائية (Chi-square) لاختبار معنوية الفروق بين القيم المشاهدة والقيم المتوقعة، وبالتالي اختبار النموذج المقدر الذي حسبت منه المشاهدات المتوقعة أي لاختبار الفرضية التالية:

H_0 النموذج مطابق لبيانات الدراسة

H_1 النموذج غير مطابق لبيانات الدراسة

جدول رقم (5) نتائج اختبار هوسمر ولمشو لاختبار معنوية النموذج

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	12.604	8	.126

يتضح من الجدول السابق أن القيمة الاحتمالية (Sig) لاختبار كاي تربيع (Chi-square) الخاصة باختبار هوسمر ولمشو أكبر من القيمة المعنوية (0.05)، وبالتالي عدم معنوية الاختبار، عليه نقبل فرضية العدم ونستنتج أن النموذج مطابق لبيانات الدراسة، كما يمكن ملاحظة ذلك من خلال جدول رقم (5) السابق حيث نلاحظ أن الاختلافات بين القيم المشاهدة والقيم المتوقعة بسيطة جداً.

جدول رقم (6) يوضح تصنيف بيانات العينة باستخدام النموذج المقدر

Classification Table ^a					
Observed			Predicted		
			هل تعتقد أن دخلك كافي لتغطية نفقات معيشة الأسرة		Percentage Correct
			غيركافي	كافي	
Step 1	هل تعتقد أن دخلك كافي لتغطية نفقات معيشة الأسرة	غيركافي	285	40	87.7
		كافي	149	71	32.3
	Overall Percentage				65.3

a. The cut value is .500

يتضح من الجدول رقم (6) السابق أن نسبة التصنيف الصحيح للدخل الغير الكافي بلغت 87.7% وذلك لبيانات عينة الدراسة باستخدام النموذج المقدر، بينما كانت نسبة التصنيف الصحيح للدخل الكافي بلغت 32.3% ربما يعزى ذلك إلى حساسية طبيعة الدراسة (ما يتعلق بكفاية الدخل) لديها الاجتماعي والمتغيرات المتشعبة التي تؤثر في ذلك، كما يتضح من الجدول أن نسبة التصنيف الصحيح للمفردات بصورة عامة باستخدام هذا النموذج بلغت 65.3% وتعتبر هذه النسبة مقبولة جداً.

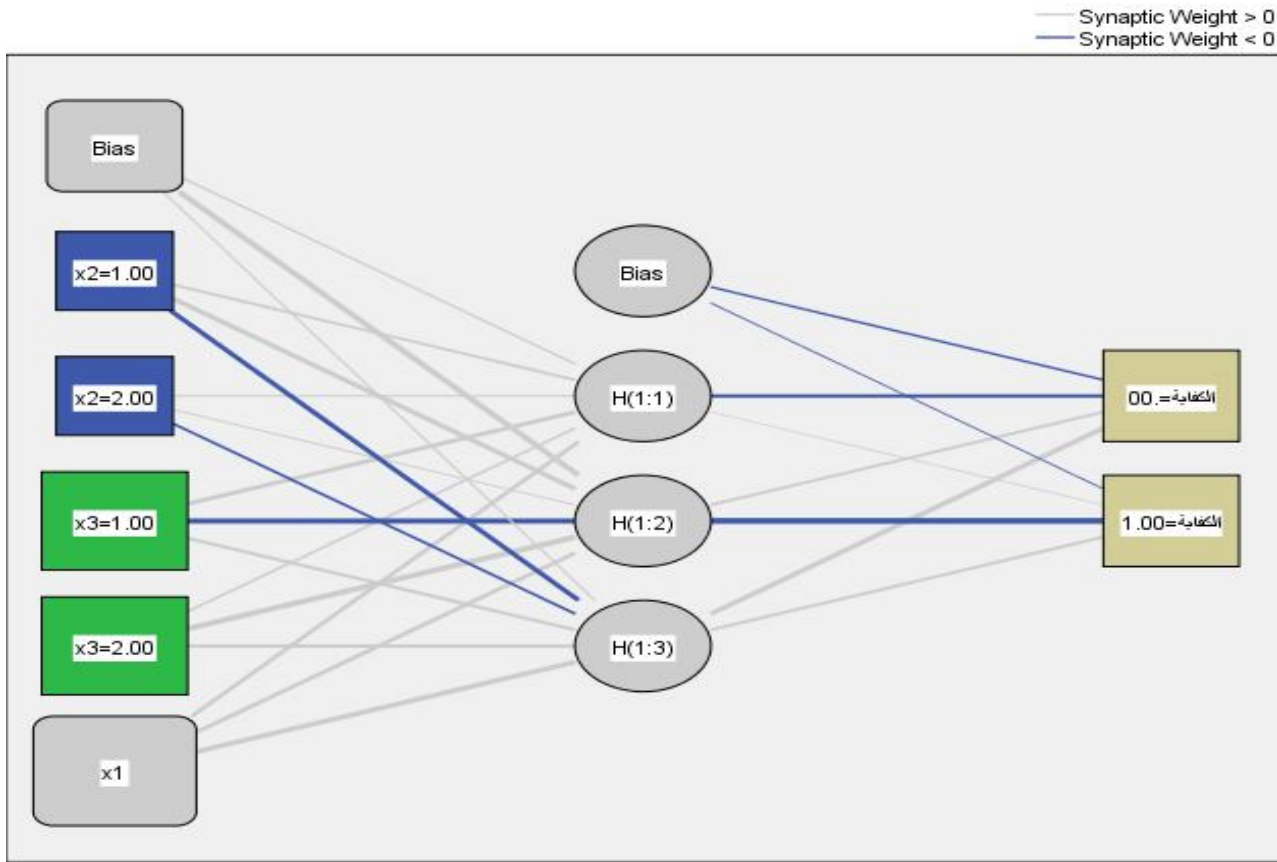
ثانياً : تطبيق اسلوب الشبكات العصبية الاصطناعية: ANN)s used Analysis :-

جدول رقم (7) معلومات الشبكة العصبية المستخدمة Network Information			
Input Layer	Factors	1	هل بالأسرة طلبة يدرسون بالجامعات
		2	طبيعة ملكية السكن
	Covariates	1	عدد أفراد الأسرة
	Number of Units ^a		5
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 ^a		3

	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	هل تعتقد أن دخلك كافي لتغطية نفقات معيشة الأسرة
	Number of Units	2	
	Activation Function	Softmax	
	Error Function	Cross-entropy	

a. Excluding the bias unit

شكل بياني رقم (3) معمارية الشبكة العصبية المستخدمة



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Softmax

جدول رقم (8) ملخص نموذج الشبكة العصبية المستخدم

Model Summary		
التدريب Training		

	نسبة التصنيف الخاطئ	34.6%
	توقف التدريب	1 consecutive step(s) with no decrease in error ^a
	زمن التدريب Training Time	0:00:00.34
الاختبار Testing		
	التصنيف الخاطئ	35.5%
Dependent Variable: هل تعتقد أن دخلك كافي لتغطية نفقات معيشة الأسرة		
a. Error computations are based on the testing sample.		

يتضح من الجدول رقم (8) السابق أن نسبة التصنيف الخاطئ في عينة التدريب بلغ 34.6% ، بينما نسبة التصنيف الخاطئ في عينة الاختبار بلغ 35.5% ، وهي نسبة متقاربة وهذا يشير إلى أن الشبكة تدرت بشكل ممتاز على تصنيف المفردات الجديدة.

جدول رقم (9) نتائج التصنيف باستخدام الشبكة العصبية المقترحة

Classification

		Percent Correct
Training	غير كافي	84.1%
	كافي	36.7%
	Overall Percent	63.6%
Testing	غير كافي	84.7%
	كافي	40.3%
	Overall Percent	69.4%

هل تعتقد أن دخلك كافي لتغطية نفقات معيشة الأسرة Dependent Variable

من الجدول رقم (9) السابق يتضح أن التصنيف الصحيح للدخل الغير كافي بلغ 84.1% في عينة التدريب بينما بلغ 84.7% في عينة الاختبار، وأن نسبة التصنيف الصحيح للدخل الكافي بلغت 36.7% في عينة التدريب بينما بلغ 40.3% في عينة الاختبار ، كما يتضح من عينة الاختبار أن نسبة التصنيف الصحيح للمفردات باستخدام هذه الشبكة العصبية الاصطناعية بلغ 69.4% ، وهي تعتبر نسبة جيدة جداً للتنبؤ بأنتماء أو تصنيف المفردات الجديدة.

جدول رقم (10) تحليل الأهمية النسبية للمتغيرات المستقلة

Independent Variable Importance

Independent Variable Importance	نسبة الأهمية في التصنيف
هل بالأسرة طلبة يدرسون بالجامعات	.083
طبيعة ملكية السكن	.373
عدد أفراد الأسرة	.543

يتضح من الجدول رقم (10) السابق أن أهم متغير مؤثر في التصنيف باستخدام الشبكة العصبية هو متغير عدد أفراد الأسرة إذ بلغت نسبة أهميته 54.3% ، ثم يليه متغير طبيعة ملكية السكن بنسبة أهمية بلغت 37.3% وأخيراً وجود طلبة يدرسون بالجامعات بنسبة أهمية 8.3% ، مما سبق يتضح أن أهم متغير لتصنيف دخل الأسرة إلى مجموعة كافي أو غير كافي هو متغير عدد أفراد الأسرة، بينما متغير وجود طلبة جامعيين بالأسرة ليس له أهمية في تصنيف دخول الأسرة أي يكاد تأثيره يساوي صفراً.

الاستنتاج:

من خلال الاطار النظري ونتائج تحليل البيانات باستخدام النموذج اللوجستي الثنائي و طريقة الشبكات العصبية توصلت الدراسة إلى النتائج التالية:

1/ إن الطريقتين لهما القدرة على التصنيف أي يمكن استخدام أي منهما لتصنيف الحالات الجديدة لدخل الأسرة إلى (كافي ، غير كافي) اعتماداً على قيم المتغيرات المستقلة لتلك الحالات.

2/ تطابق نتائج الطريقتين من حيث أهمية المتغيرات المستقلة المؤثرة معنوياً وغير المؤثرة في عملية تصنيف المفردات الجديدة ، حيث اثبتت كلتا الطريقتين أن حجم الأسرة يعتبر أهم متغير لتحديد كفاية دخل الأسرة بينما متغير وجود طلبة يدرسون بالجامعات ليس له تأثير في التصنيف.

3/ من الجدول رقم (7) يتضح أن التصنيف الصحيح باستخدام النموذج اللوجستي الثنائي المقدر بلغ 65.3%، بينما يتضح من الجدول رقم (10) أن نسبة التنبؤ أو التصنيف الصحيح باستخدام نموذج الشبكات العصبية بلغت 69.4%، مما سبق نستنتج أن استخدام الشبكات العصبية أفضل من استخدام النموذج اللوجستي الثنائي.

4/ يعتبر حجم الأسرة أهم متغير مؤثر على كفاية دخل الأسرة ، وأن وجود طلبة يدرسون بالجامعات ليس له تأثير يذكر على كفاية دخل الأسرة.

التوصيات:

بناءً على ما سبق من نتائج نوصي بالآتي:

1/ الاستفادة من الأساليب الإحصائية المتقدمة مثل النموذج اللوجستي، ونماذج التصنيف الحديثة المتمثلة في نماذج الشبكات العصبية للفصل أو التمييز بين مجموعتين أو أكثر، في جميع مجالات المعرفة إذا كانت المتغيرات خليط بين النسب و المتقطعة أو لا تتبع التوزيع الطبيعي.

2/ تعميم فكرة استخدام النموذج اللوجستي الثنائي وأساليب الشبكات العصبية في المجالات الاجتماعية والاقتصادية وعدم تركيزها على المجالات الطبية فقط كما كان في السابق.

3/ إذا كان لدينا نموذج ذو متغير تابع نوعي (ثنائي الاستجابة)، وخليط (كمية ، نوعية) من المتغيرات المستقلة، أو إذا كان المتغير التابع أو بعض المتغيرات المستقلة لا تتبع التوزيع الطبيعي يفضل استخدام أسلوب الشبكات العصبية الاصطناعية أو النموذج اللوجستي الثنائي بدلاً عن الدالة التمييزية.

4/ إجراء مزيد من الدراسات على الدخل الأسري و المتغيرات المؤثرة فيه ، لمعرفة أكثر المتغيرات تأثيراً على كفاية الدخل، ذلك لوضع التدابير و الحلول اللازمة ، لان عدم كفاية الدخل يؤثر بصورة مباشرة على نوعية صحة وتعليم الأسرة ، وجميع النواحي الحياتية الأخرى ، كذلك يمكن أن يؤثر على حجم الجريمة و مستوى الأمن القومي.

المراجع و المصادر:-

1- المرجع و المصادر العربية:

- 1) ريتشارد جونسون & دين وشرن (1998م) ، تعريب : عبد المرضي حامد عزام ،
- 2) التحليل الإحصائي متعدد المتغيرات دار المريخ للنشر ، الرياض ، المملكة العربية السعودية ، ص: 735-736 ، 822 - 823.
- 3) طاهر ريسان دخيل (2008م) ، استخدام الشبكات العصبية لأغراض التمييز، ورقة بحثية منشورة، مجلة العلوم الاقتصادية و الإدارية، المجلد (14)، العدد (52) ، ص: 1-7.

4) عباس ناجي جواد (2010م) ، المفاضلة بين طرق تقدير الدوال الاقتصادية ذات المتغيرات التابعة. النوعية، ورقة عمل منشورة، مجلة تكريت للعلوم الإدارية والاقتصادية، المجلد (6)، العدد (18) ، ص: 2-4.

5) عبد الحميد محمد العباسي، 2003م ، مقدمة في الشبكات العصبية الاصطناعية وتطبيقاتها في العلوم الاجتماعية باستخدام SPSS، معهد الدراسات و البحوث الإحصائية، جامعة القاهرة، ص: 1 - 8.

6) عبد المرضي عزام (1998م) ، التحليل الإحصائي للمتغيرات المتعددة، دارالمريخ للطباعة والنشر، الرياض.

7) محمد عبد الرحيم حمدي وكرم حاتم ذنون (2009م) ، تشخيص مرض التدنن الرئوي (السل) باستخدام الشبكات العصبية الاصطناعية ، ورقة بحثية منشورة، مجلة الرافيدين لعلوم الحاسبات و الراضيات ، المجلد (6)، العدد (1) ، ص: 2-3.

8) محمود خالد عكاشة(2002م)، استخدام نظام SPSS في تحليل البيانات الإحصائية، الطبعة الاولى ، مطبعة جامعة الأزهر - غزة - فلسطين ، ص: 492.

9) وفاء مصطفى سيد أحمد (2010م)، استخدام الدالة التمييزية لدراسة مستوى الإصابة بمرض الفشل الكلوي، رسالة ماجستير غير منشورة، مكتبة كلية العلوم - جامعة السودان للعلوم والتكنولوجيا، الخرطوم ، ص: 24 - 42.

2- المراجع و المصادر الإنجليزية:

- 1- P. Dixon, W. M. Duckworth, M. S. Kaiser, K. Koehler, W. Q. Meeker and W.R. Stephenson,(2007), Binary Response and Logistic Regression Analysis.
- 2- O'Connell, Ann A, (2006), Logistic regression models for ordinal response variables, C&M Digital (P) Ltd., India - New Delhi.