المقارنة بين نموذج الانحدار اللوجستي والشبكات العصبية في تصنيف المشاهدات (دراسة ميدانية: تحديد أهم العوامل المؤثرة على كفاية دخل الأسرة في مدينة القاهرة)

د. عماد الدين ابراهيم علي[1]

الملخص:

تناولت هذه الدراسة المقارنة بين نموذج الانحدار اللوجستي Logistic Regression Model وأسلوب الشبكات العصبية الاصطناعية Artificial Neural Networks في القدرة على تصنيف المشاهدات بالتطبيق على دراسة أهم العوامل المؤثرة على كفاية دخل الاسرة في مدينة القاهرة.

الجدير بالذكران نموذج الانحدار اللوجستي من أكثر النماذج ملائمة لتحليل المتغير التابع الوصفي الثنائي والمتعدد، وهو يمثل حالة خاصة من نموذج الانحدار الخطى العام، لكنه مقيد في تطبيقه، فاذا كانت العلاقة بين المتغيرات المستقلة غير خطية فإن النتائج المتحصل عليها من الانحدار اللوجستي تكون مضللة، أما الشبكات العصبية فهي أسلوب لا يشترط في استخدامه على ان تكون العلاقة خطية أو غير خطية، كما انه لا يشترط توزيع احتمالي معين (الطبيعي) لمتغيرات الدراسة، لذا يعتبر اسلوب الشبكات العصبية أكثر ملائمة في الاستخدام في مثل هذه الظروف من نماذج التصنيف الأخرى.

في هذه الدراسة نقدم مقارنة عملية لقياس القدرة على التصنيف بين نموذج الانحدار اللوجستي والشبكات العصبية الاصطناعية.

الكلمات المفتاحية:

نموذج الانحدار اللوجستي – الشبكات العصبية الاصطناعية – دخل الأسرة

Abstract:

This study compares the Logistic regression model and the Artificial neural networks in terms of their ability to classifying observations related to a study about the most important factors affecting household income in the city of Cairo. It is worth mentioning that, the logistic regression model is one of statistical models that used when the dependent variable is a dichotomous or polychotomous. It is a special case of linear regression model, and hence restrictively relevant in the sense that results obtained from it may be useless if linearity does not hold.

On the other hand, "Artificial Neural networks" is a method of analysis based on both linear and non-linear relationships, which makes it more relevant in such circumstances. This study presents a real life application of these two methods in order to compare the performance of the two models.

Keywords:

Logistic regression model, Artificial Neural Networks, Household income

[1] مدرس بكلية التجارة – جامعة عين شمس

مقدمة:

تعد عملية التصنيف بين المشاهدات من الأساليب الشائعة الاستخدام، ويرجع ذلك الى كثرة الظواهر التي يمكن تحليلها وتفسيرها من خلال أساليب التصنيف المختلفة. وقد تعددت الأساليب والطرق المستخدمة في عملية التصنيف.

وتعد الشبكات العصبية Artificial Neural Networks أحد الاساليب الاحصائية المستخدمة في عملية التصنيف والفصل بين المشاهدات، كما تستخدم في جوانب تطبيقية أخرى مثل التنبؤ بالسلوك الاستثماري والكشف عن الظواهر الطبية ومحاكاة عمل الرجل الألى وتحليل الصورة، وذلك للمرونة التي يتمتع بها اسلوب الشبكات العصبية وقدرته على التعامل مع الدوال غير الخطية، لأنه لا يعتمد على نوع التوزيع الاحتمالي الذي تتبعه متغيرات الظاهرة محل الدراسة.

وخلال هذه الدراسة يحاول الباحث اجراء مقارنة بين نموذج الانحدار اللوجستي وأسلوب الشبكات العصبية، وذلك لتصنيف المشاهدات الى المجموعة التي تنتمي اليها في حال كون بعض المتغيرات لا تتبع التوزيع الطبيعي.

تناول الجانب التطبيقي دراسة أهم العوامل المؤثرة على كفاية دخل الاسرة في مدينة القاهرة، وقد حظيت الاسرة باهتمام وعناية الباحثين في مختلف العلوم الاقتصادية والاجتماعية، ويأتي هذا الاهتمام لما تحتله الاسرة من أهمية كبيرة في منظومة العلاقات الاقتصادية والاجتماعية.

مشكلة الدراسة:

مع تعدد الأساليب والطرق الإحصائية المستخدمة في عملية التصنيف والفصل بين المشاهدات، كان لابد من تحديد أفضل اسلوب يمكن استخدامه في تصنيف المشاهدات بحيث يحقق أعلى دقة تصنيف ممكن مع الاخذ في الاعتبار بشروط تطبيق كل اسلوب وفروضه على البيانات محل الدراسة. وسوف يتم الاعتماد على نسبة المشاهدات المصنفة بشكل صحيح (نسبة التصنيف الصحيح) كمعيار للمقارنة.

أهداف الدراسة:

- التعرف على مفهوم الشبكات العصبية الاصطناعية كأحد اساليب التصنيف الحديثة.
- تحديد الأسلوب الإحصائي الذي يحقق أعلى دقة تصنيف ممكنة وذلك من خلال المفاضلة بين نموذج الانحدار اللوجستي واسلوب الشبكات العصبية.
- تحديد أهم العوامل التي تؤثر معنوياً على كفاية دخل الاسرة في مدينة القاهرة.

أهمية الدراسة:

- الحاجة إلى استخدام أحد أساليب التصنيف الحديثة والتي لا تتطلب شروط معينة للاستخدام مثل أسلوب الشبكات العصبية الاصطناعية.
- تطبيق نموذج الانحدار اللوجستي وكذلك أسلوب الشبكات العصبية على بعض التطبيقات الاقتصادية (كفاية دخل الاسرة) يعطى اهمية للدراسة، خاصة أن هذه النماذج شائعة الاستخدام في الدراسات الطبية والاجتماعية.
- دراسة اهم العوامل المؤثرة على كفاية دخل الاسرة في مدينة القاهرة من وجهة نظر إحصائية مع تحديد مدى تأثير ومعنوية تلك العوامل يساهم في وضع خطط التنمية الشاملة وتحقيق الاستقرار الاقتصادي والاجتماعي للأسرة.

الدراسات السابقة:

الدراسات العربية:

- دراسة هلا بسام عبد الله (2004)

في هذه الدراسة تم استخدام النموذج اللوجستي للتوصل الى اهم النسب المالية التي يمكن استخدامها في التمييز بين الشركات المتعثرة والشركات غير المتعثرة، وتم التطبيق عل قطاع المقاولات، اتفقت هذه الدراسة مع الدراسة الحالية في استخدام نموذج الانحدار اللوجستي في عملية التصنيف ولكنها لم تقارن بين أكثر من طريقة في التصنيف.

- دراسة عدنان غانم وفريد الجاعونى (2007)

تناولت هذه الدراسة التحليل الإحصائي متعدد المتغيرات لتوزيع الأسر داخل الهيكل الاقتصادي والاجتماعي داخل المجتمع، وركزت الدراسة على تحليل التمايز في عملية التصنيف، بينما تركز الدراسة الحالية على نموذج الانحدار اللوجستي والشبكات العصبية.

- دراسة عباس ناجى جواد (2010)

خلال هذه الدراسة تم المفاضلة بين طرق تقدير الدوال الاقتصادية ذات المتغيرات التابعة النوعية، واثبات فرضية الدراسة التي تنص على أنه لا يمكن استخدام طريقة المربعات الصغرى العادية في تقدير نماذج الانحدار ذات المتغيرات التابعة النوعية، ومن ثم اثبات أن تحليل التمايز أو الانحدار اللوجستي أو نموذج لوجيت أو بروبت يؤدى الى تقديرات ادق وأكثر منطقية ولم تتناول الدراسة الشبكات العصبية.

- دراسة عدنان غانم وفريد الجاعونى (2011)

تناولت الدراسة استخدام نموذج الانحدار اللوجستي ثنائي الاستجابة في دراسة أهم المحددات الاقتصادية والاجتماعية لكفاية دخل الاسرة، واستهدفت الدراسة تحديد أهم المحددات المؤثرة معنوياً على كفاية دخل الاسرة، ولكن لم تتطرق الدراسة الى مقارنة نتائج الانحدار اللوجستي بنتائج نماذج أخرى.

- دراسة ادم الطاهر نوح (2015)

تناولت الدراسة استخدام نوعين من التحليل هما تحليل التمايز والشبكات العصبية لتصنيف وتحليل فئات الدخل في السودان، وتوصلت الدراسة الى أن دالة التمايز الناتجة من تحليل التمايز كانت معنوية ولكن كفاءة الدالة في التمييز لم تكن عالية، أما اسلوب الشبكات العصبية فقد أعطى دالة تمايز بمتوسط مربع خطأ أكبر من متوسط مربع الخطأ في تحليل التمايز، واوصت الدراسة باستخدام شبكات اخرى في عملية التصنيف لتحسين مستوى الخطأ.

- دراسة فادية أدم على (2016)

استهدفت الدراسة بناء نموذج رياضي يمكن من خلاله التمييز بين كفاية دخل رب الأسرة بولاية شمال كردفان باستخدام اسلوب تحليل التمايز وتحديد اهم العوامل الاقتصادية والاجتماعية التي تؤثر على كفاية دخل رب الاسرة.

الدراسات الاجنبية:

- دراسة Tang, Man-Lai (2001)

استهدفت هذه الدراسة تحديد أدق اختبار جودة توفيق للنموذج اللوجستي الثنائي، وأهتمت الدراسة بمدى تأثير التحصيل الأكاديمي على تقييم الذات باستخدام النموذج اللوجستي الثنائي، واجريت الدراسة على مجموعتين من الطلاب (البيض والسود)، وقد تم التوصل إلى انه لا يوجد فرق معنوي بين الطلاب البيض والسود من حيث تأثير التحصيل الأكاديمي على تقييم الذات، واتفقت هذه الدراسة مع الدراسة الحالية في استخدام نموذج الانحدار اللوجستي واختلفت في مجال التطبيق.

- دراسة Rahman, Atta (2009)

في هذه الدراسة تم تحديد أهم العوامل الاجتماعية والاقتصادية التي تؤثر على التعليم في منطقة NWFP في دولة باكستان، وتم استخدام نموذج الانحدار اللوجستي لتحليل بيانات الدراسة، وتم التعبير عن المتغير التابع بمتغير ثنائي (الشخص متعلم – الشخص غير متعلم) وتمثلت المتغيرات المستقلة في مستوى تعليم الأب، دخل الأب، موقف الأبوين تجاه التعليم، مستوى تعليم الأم، النظام الحالي للامتحانات، والمنهج التعليمي الحالي، وتوصلت الدراسة إلى تحديد أهم العوامل الاجتماعية والاقتصادية التي تؤثر على التعليم في تلك المنطقة، وتتفق هذه الدراسة مع الدراسة الحالية في استخدام نموذج الانحدار اللوجستي في التصنيف ولم تتطرق إلى المقارنة بطرق تصنيف اخرى.

- دراسة Thomas.N,Anne Wangombe and Nancy Khadioli (2010)

تناولت الدراسة استخدام الانحدار اللوجستي لتحديد اهم العوامل المحددة للفقر في دولة كينيا باستخدام بيانات المسح السكاني والصحة، وتمثل المتغير التابع في الحالة الاقتصادية والاجتماعية للأسرة (فقيرة – غير فقيرة)، واعتبار العوامل الديموجرافية هي المتغيرات المستقلة، واظهرت النتائج أن الزيادة في التعليم له تأثير كبير على تخفيض احتمال ان تكون الاسرة فقيرة، وان الأسرة الريفية لها احتمال اكبر في ان تكون فقيرة، واثبتت الدراسة ايضا أن مستوى دخل الأسرة واحتمال كونها فقيرة يتأثر بعمر رب الأسرة والديانة والمنطقة السكنية والعامل العرقي.

اتفقت هذه الدراسة مع الدراسة الحالية في استخدام نموذج الانحدار اللوجستي الثنائي، وكذلك مجال التطبيق وهو المجال الاقتصادي، ووجه الاختلاف أن هذه الدراسة لم تتطرق إلى المقارنة بطرق تصنيف اخرى.

بعد استعراض الدراسات السابقة يلاحظ وجود اتفاق بين هذه الدراسات والدراسة الحالية في التركيز على دخل الاسرة والعوامل المؤثرة فيه، بينما تختلف في نوع الاسلوب المستخدم في عملية التصنيف وعليه سوف تكون هذه الدراسة مكملة للدراسات السابقة حيث تهدف الى استخدام نموذج الانحدار اللوجستي وكذلك أسلوب الشبكات العصبية في تصنيف المشاهدات ودراسة اهم العوامل المؤثرة على كفاية دخل الاسرة في مدينة القاهرة.

فروض الدراسة:

في ضوء مشكلة الدراسة وبعد استعراض نتائج الدراسات السابقة يمكن صياغة فروض الدراسة على النحو التالي:

1. استخدام الشبكات العصبية الاصطناعية في عملية التصنيف بين المشاهدات يحقق كفاءة عالية.
2. الاعتماد على استخدام اسلوب الشبكات العصبية في عملية التصنيف عندما تكون المتغيرات المستقلة لا تتبع التوزيع الطبيعي يحقق نسبة تصنيف عالية عن اسلوب الانحدار اللوجستي.
3. لا توجد علاقة ذات دلالة إحصائية بين كفاية دخل الاسرة كمتغير تابع وكل من المتغيرات المستقلة: عدد أفراد الاسرة، متوسط الدخل الشهري للأسرة، الحالة الوظيفية لرب الاسرة، المستوى التعليمي لرب الأسرة، نوعية السكن، وجود طلبة جامعيين بالأسرة، واخيراً وجود أفراد بالأسرة لديهم مرض مزمن.

حدود الدراسة:

- الحدود المكانية: مدينة القاهرة – مصر
- الحدود الزمانية: تبدأ من يناير 2016 إلى ديسمبر 2016

مصطلحات الدراسة:

دخل الأسرة: هو الدخل الإجمالي لجميع أفراد الأسرة نقداً أو عيناً أو خدمات سنوياً أو على فترات أقل، ويشمل ذلك كل شكل من اشكال الدخل بما في ذلك الرواتب والأجور ودخل التقاعد والتحويلات الحكومية النقدية والمكاسب الاستثمارية، ويعتبر دخل الاسرة معيار لتحديد مستوى معيشة الأسرة.

الدراسة الميدانية:

تعد الدراسة الميدانية المصدر الرئيس للبيانات الاولية للدراسة، حيث تم الحصول على بيانات متغيرات الدراسة من خلال المقابلات الشخصية مع أرباب أو ربات الأسر التي يقع عليها الاختيار ضمن العينة. وتم جمع البيانات[2] واجراء المقابلات بمعرفة الباحث كباحث رئيسي وبمعاونة فريق بحثي ضم مجموعة من المعيدين وطلاب الدراسات العليا.

- مجتمع وعينة الدراسة:

يتمثل مجتمع الدراسة في جميع الاسر داخل الاحياء الاربعة (عين شمس – الوايلى – الزيتون – الخليفة) التي تم اختيارها عشوائياً من مناطق مدينة القاهرة الأربعة (الشرقية – الغربية – الشمالية – الجنوبية) وقد بلغ مجموع الاسر في الاحياء الاربعة 273213[3] أسرة، وتم تحديد حجم العينة عند

[2] تم جمع البيانات خلال الفترة من 2016/1/1 إلى 2016/6/31

[3] الجهاز المركزي للتعبئة العامة والإحصاء – الكتاب الإحصائي السنوي 2016

مستوى معنوية 5% ، ودرجة ثقة 95% ليصل إلى 384[4] اسرة وتم زيادة حجم العينة بمقدار 77 أسرة بنسبة 20% من حجم العينة فبلغ حجم العينة الكلى 461 أسرة وذلك تحسباً لعدم الاستجابة أو عدم تواجد بعض الأسر اثناء المقابلات، وبلغت نسبة الاستجابة من قبل الأسر محل الدراسة 89.8% أي ما يعادل 414 أسرة، تم تخصيص 30 اسرة منهم كعينة استطلاعية و384 اسرة كعينة اساسية للدراسة وتم توزيع عينة الدراسة 384 توزيعاً نسبياً على الاحياء الأربعة حسب عدد الأسر داخل كل حي وكانت على النحو التالي:

جدول رقم (1)
توزيع حجم العينة على الاحياء الاربعة

الحى	عدد الأسر داخل كل حي	التخصيص
عين شمس	153000	$n_1 = \frac{153000}{273213} X384 = 215$
الوايلى	36884	$n_2 = \frac{36884}{273213} X384 = 52$
الزيتون	43714	$n_3 = \frac{43714}{273213} X384 = 61$
الخليفة	39615	$n_4 = \frac{39615}{273213} X384 = 56$
الإجمالي	273213	$n = 384$

المصدر: من اعداد الباحث استنادا إلى بيانات الجهاز المركزي للتعبئة العامة والإحصاء عام 2016

ويوضح الجدول رقم (2) توزيع أفراد العينة وفقاً للنوع والعمر والحالة الاجتماعية

جدول رقم (2)
توصيف عينة الدراسة وفقاً للنوع والعمر والحالة الاجتماعية

الاحياء	النوع		العمر				الحالة الاجتماعية			
	ذكور	إناث	أقل من 20	30-20	40-30	40 فاكثر	اعزب	متزوج	مطلق	ارمل
عين شمس	168	47	10	56	60	89	11	143	36	25
الوايلى	41	11	2	11	16	23	3	35	9	5
الزيتون	47	14	4	10	22	25	4	36	12	9
الخليفة	44	12	2	9	20	25	3	35	12	6
الإجمالي	300	84	18	86	118	162	21	249	69	45
النسب المئوية	%78.1	%21.9	%4.7	%22.4	%30.7	%42.2	%5.5	%64.8	%17.9	%11.8

المصدر: من إعداد الباحث

[4] تم حساب حجم العينة وفقاً للمعادلة التالية:

$$n = \frac{z^2 p(1-p)}{\epsilon^2} = \frac{(1.96^2)(0.5)(0.5)}{(0.05^2)} = 384$$

- تصميم استمارة الاستقصاء:

تم التصميم المبدئي للاستمارة من خلال الدراسات السابقة وخبرة الباحث، وتم تعديل الاستمارة بناء على عينة استطلاعية من (30) أسرة بغرض استطلاع الآراء حول إمكانية إضافة نقاط أوحذف أو تعديل بعض العبارات لعدم وضوحها، وقد استخدم الباحث أسلوب المقابلة الشخصية خلال هذه المرحلة، وأسفر ذلك عن الشكل الحالي للاستمارة والتي تحتوي على جزئين، الجزء الأول يحتوي على بيانات عامة عن المستقصي منه، والجزء الثاني يحتوي على ثمانية أسئلة خاصة بأهم المتغيرات المؤثرة على كفاية دخل الأسرة.

- متغيرات الدراسة:

شملت الدراسة على متغير تابع واحد وسبع متغيرات مستقلة وكانت على النحو التالي:

المتغير التابع (Y)

يعبر المتغير التابع عن كفاية دخل الأسرة وهو متغير ثنائي يأخذ القيمة (0) إذا كان دخل الأسرة غير كافي، والقيمة (1)إذا كان دخل الأسرة كافي.

المتغيرات المستقلة (X'S)

حيث ان الهدف الرئيسي للدراسة هو المفاضلة بين استخدام نموذج الانحدار اللوجستي وأسلوب الشبكات العصبية في تصنيف المشاهدات، فقد تمثلت المتغيرات المستقلة في سبعة متغيرات فقط من بين المتغيرات التي يمكن أن تؤثر على المتغير التابع (كفاية دخل الأسرة) والمتغيرات المختارة هي من الواقع المشاهد في الحياة العملية واراء الكثير من الاقتصاديين، وهي:

X1	عدد أفراد الأسرة (حجم الاسرة)[5]	4 أفراد فأقل = 0	أكثر من 4 افراد = 1
X2	متوسط الدخل الشهري للأسرة[6]	4000 جنيه فأقل = 0	اكثر من 4000 جنيه = 1
X3	الحالة الوظيفية لرب الأسرة	لا يعمل = 0	يعمل = 1
X4	المستوى التعليمي لرب الأسرة	غير متعلم = 0	متعلم = 1
X5	نوعية السكن	إيجار = 0	ملك = 1
X6	وجود طلبة جامعيين بالأسرة	لا يوجد = 0	يوجد = 1
X7	وجود أفراد بالأسرة لديهم مرض مزمن	لا يوجد = 0	يوجد = 1

النماذج الإحصائية المستخدمة في الدراسة:

اعتمدت الدراسة على استخدام:

اولاً: الشبكات العصبية الاصطناعية ثانياً: نموذج الانحدار اللوجستي

وفيما يلي عرضاً مبسطاً للخلفية النظرية لكل أسلوب على حدة

أولاً: الشبكات العصبية الاصطناعية Artificial Neural Networks (ANN)

تعتبر الشبكات العصبية أحد مجالات الذكاء الاصطناعي، وهي عبارة عن صيغ رياضية تعتمد على نماذج رياضية تحاكى عمل المخ البشرى في حل المشكلات واجراء الحسابات، ويطلق على الشبكات العصبية عدة مسميات اخرى مثل أنظمة التوزيع المتوازي أو الانظمة المترابطة أو الأنظمة التكيفية

(1) مكونات الشبكة العصبية الاصطناعية:

تتكون أي شبكة عصبية اصطناعية من مجموعة من وحدات المعالجة تسمى نيورون (Neuron) وهذه الوحدات متصلة فيما بينها في شكل ترابطات تسمى بالشبكة العصبية، حيث أن مخرجات أي نيورون تعتبر مدخلات لنيورون اخر، ويتم تقسيم النيورونات إلى مجموعات كل مجموعة تسمى طبقة وهناك ثلاث أنواع من الطبقات بالإضافة الى الوصلات البينية والطبقات هي:

- طبقة المدخلات (Input Layer) وهي الطبقة التي من خلالها يتم إمداد الشبكة بالبيانات.

---

(5) متوسط عدد افراد الأسرة عام 2016 = 3.64 ≈ 4 أفراد

(6) متوسط الدخل الشهري للأسرة عام 2016 = 3683.33 ≈ 4000 جنية

المصدر: الجهاز المركزي للتعبئة العامة والإحصاء - نتائج بحث الدخل والأنفاق والاستهلاك (2016)

- الطبقة الخفية (Hidden Layer) ويتم خلالها معالجة البيانات ثم أرسالها إلى طبقة المخرجات.
- طبقة المخرجات (Output Layer) وخلال هذه الطبقة يتم الحصول على مخرجات الشبكة أي الناتج او حل للمشكلة محل الدراسة.
- الوصلات البينية (Connections) تمثل وصلات اتصال بين طبقات الشبكة المختلفة أو بين الوحدات داخل كل طبقة عبر الأوزان والتي تمثل قوة الاتصال بين النيورونات، فاذا كان لدينا النيورون (i) يتصل بالنيورون (j) فسوف نرمز لوزن الاتصال بينهم بالرمز (w_{ij}) (عبدالعال ،2004)

(2) أنواع الشبكات العصبية الاصطناعية:

يمكن تصنيف الشبكات العصبية الاصطناعية حسب طبيعة انتشار البيانات خلال طبقات الشبكة إلى شبكات التغذية الأمامية (Feed Forward Networks)، وشبكات التغذية الخلفية (Feed Back Networks)، وشبكات ذاتية التنظيم (Self-Organizing Networks). (عيسى ،2000)

(3) التصميم الهيكلي لبناء الشبكة العصبية الاصطناعية:(العباسي ،2013)

يتضمن التصميم الهيكلي لبناء الشبكات العصبية الاصطناعية (ANN) الخطوات التالية:

- تجميع البيانات التي تستخدم في تدريب او اختيار الشبكة.
- تعريف البيانات الخاصة بالتدريب للشبكة، ووضع خطة التدريب والتعلم.
- بناء هيكل الشبكة وتحديد نوع الشبكة ومكوناتها من حيث عدد الطبقات.
- اختيار طريقة التعلم حسب ما هو متاح من أدوات تطوير الشبكة.
- تحديد قيم للأوزان والمتغيرات، ثم تعديل قيم الأوزان عن طريق التغذية العكسية.
- تحويل البيانات إلى النوع المناسب للشبكة من خلال المعادلات.
- اجراء عمليتي التدريب والاختبار من خلال تكرار عرض المدخلات والمخرجات المرغوبة إلى الشبكة، ومنها مقارنة القيم الفعلية مع القيم المحسوبة، ثم حساب الخطأ، ثم تعديل الأوزان لتخفيض الخطأ حتى يصبح مقبولاً.
- الوصول إلى النتائج المستهدفة من خلال استخدام مدخلات التدريب، وبالتالي يمكن الاعتماد على الشبكة في الاستخدام كنظام مستقل قائم بذاته أو كجزء من النظام.

(4) العناصر الأساسية اللازمة لبناء الشبكة العصبية:

(أ) البيانات

تقسم البيانات (المشاهدات) إلى ثلاث فئات هي فئة التدريب وفئة الصلاحية واخيراً فئة الاختبار.

(ب) دوال المدخلات ودوال التحويل

- دوال المدخلات (دالة الجمع): (Summation Function)

أن أول عملية يقوم بها النيورون هي حساب قيمة المدخلات باستخدام دالة الجمع التالية:

$$S_j = \sum_{i=1}^{n} x_i w_{ij} \qquad (1)$$

حيث:

S_j : ناتج عملية الجمع للنيورون (j)

x_i : المخرجات القادمة من النيورون (i) ومتجه إلى النيورون (j)

w_{ij} : الوزن الذي يربط النيورون (j) بالنيورون (i)

- دوال التحويل: (Transformation Function)

بعد عملية الجمع تبدأ عملية تحويل ناتج الجمع إلى أحد القيم التي يفترض ان تكون ضمن نواتج الشبكة المرغوب فيها وتتم هذه الخطوة باستخدام دالة تسمى دالة التحويل. وتعتبر الدالة اللوجستية من أكثر دوال التحويل استخداماً، حيث تكون المخرجات أرقاماً محصورة بين الصفر والواحد الصحيح، وتأخذ الشكل التالي:

$$F(x) = \frac{1}{1+e^{-s}} \qquad (2)$$

$$S = \sum_{i=1}^{n} x_i w_i + \theta \qquad (3)$$

وهناك دوال أخرى للتحويل مثل الدالة الخطية، دالة الخطوة، ودالة الاشارة.

(جـ) طرق التعلم (طرق التدريب):

طرق التعلم تستخدم في اعطاء الشبكة القدرة على التعلم حتى الوصول إلى المخرجات المستهدفة بأقل خطأ، وهناك نوعان من التعلم هما: التعلم الذاتي والتعلم غير الموجه.

(د) معلمات التعلم:

يقصد بمعلمات التعلم الأدوات التي تستخدم في تحسين اداء الشبكة العصبية وهناك ثلاث معلمات للتعلم وهي

- معدل التعلم: (Learning Rate)

وهو يمثل حدود تعديل الأوزان وكلما زاد معدل التعلم زادت قدرة الشبكة على التعلم.

- عامل الدفع: (Momentum Factor)

يمثل عامل الدفع نسبة التحيز في الأوزان من مرحلة إلى اخرى، وللحصول على شبكة عصبية ثابتة يجب ان يكون هذا العامل أقل من الواحد الصحيح.

- حد التجاوز عن خطأ التدريب: (Training Tolerance)

قيمة هذا الحد تمثل الخطأ المسموح به اثناء المقارنة بين مخرجات الشبكة والمخرجات الحقيقية، وإذا بلغت قيمة هذا الحد الصفر فهذا يعنى أن مخرجات الشبكة مطابقة للمخرجات الحقيقية، وإذا زادت القيمة فإن هذا يعنى انخفاض دقة التنبؤات، ويحدد هذا العامل بالمحاولة والخطأ واعتماداً على خبرة الباحث وطبيعة البيانات.

(5) خصائص الشبكات العصبية الاصطناعية:

من أهم خصائص الشبكات العصبية الاصطناعية ما يلي:

- تطبق الشبكات العصبية بغض النظر إلى تحقق فرضيات معينة عن طبيعة المتغيرات المستخدمة في التحليل وطبيعة علاقاتها مع بعضها البعض، وهذا يعطى مبرراً لاستخدام الشبكات العصبية دون الطرق الإحصائية التقليدية الاخرى. (عبد العال ،2004)
- امكانية تطبيق الشبكات العصبية في العديد من المجالات العلمية مثل المجالات الاقتصادية والمالية وكذلك مجال سوق المال والاستثمار كالتنبؤ بالفشل المالي والتنبؤ بمعدل أسعار الاسهم والسندات ومخاطر الإقراض وغيرها وكذلك استخدامها في المجالات الحيوية وايضا المجالات الطبية.
- يمكن استخدام الشبكات العصبية كأداة للتحليل بدلاً من الاساليب الإحصائية التالية: التصنيف (التمييز)، التحليل العنقودي، التنبؤ
- تبنى الشبكات العصبية على أساس رياضي قوى وتتعامل مع البيانات الكمية والنوعية.
- مفهوم الاستدلال الإحصائي يتضح في نماذج الشبكات العصبية من خلال عملية التدريب او التعلم.

ثانياً: نموذج الانحدار اللوجستي الثنائي Logistic Regression Model Binary

يستخدم الانحدار اللوجستي عندما يكون المتغير التابع (Y) ثنائي الانقسام، أي يأخذ القيمة (1) في حالة وقوع الحدث محل الاهتمام وذلك باحتمال قدره (P) بينما يأخذ القيمة (0) في حالة عدم وقوع الحدث محل الاهتمام باحتمال قدره (1- P)، كذلك لا يضع الانحدار اللوجستي قيوداً على أنواع المتغيرات المستقلة (X's) والتي يمكن لها أن تكون متصلة أو فئوية أو خليط من الاثنين، (Lea, 1997;Pample, 2000; King, 2003)

من المعلوم أن معادلة الانحدار الخطى البسيط تكون على الصورة:

$$y|x = B_0 + B_1 x + e \qquad (4)$$

حيث يعنى $y|x$ المتغير التابع Y بشرط حدوث المتغير المستقل X ، وبافتراض أن الخطأ العشوائي (e) يتبع التوزيع الطبيعي بمتوسط (0) وانحراف معياري $\sigma_{y/x}$ أي أن($e \sim N(0, \sigma_{y/x})$ فان المتغير التابع Y يتبع التوزيع الطبيعي بمتوسط $\mu_{y/x}$ وانحراف معياري $\sigma_{y/x}$ أي أن ($Y \sim (\mu_{y/x}, \sigma_{y/x})$ وذلك لكل قيمة من قيم المتغير المستقل X .

ونظراً لان$E(e)=0$ لذا فإن القيمة المتوقعة للمتغير Y عند قيمة معينة للمتغير X تكون على الشكل التالي:

$$E(y|x) = \hat{B}_0 + \hat{B}_1 x \qquad (5)$$

الا انه ولعدم إمكانية تطبيق الانحدار الخطى البسيط في حالة كون المتغير التابع وصفياً لذلك يتم استخدام النموذج اللوجستي الذي يعالج المشكلة السابقة، حيث يمكن كتابته في حالة وجود متغير مستقل واحد كما يلي:

$$\log_e \left(\frac{p}{1-p}\right) = \hat{B}_0 + \hat{B}_1 x \qquad (6)$$

$$wher: E(e) = 0$$

وبصورة أخرى

$$\frac{p}{1-p} = e^{\hat{B}_0 + \hat{B}_1 x}$$

حيث:

p : احتمال وقوع الحدث محل الاهتمام أي احتمال النجاح

$p-1$: احتمال عدم وقوع الحدث أي احتمال الفشل

$\left(\frac{p}{1-p}\right)$: نسبة الترجيح للحدث محل الاهتمام (odds)

$\log_e \left[\frac{p}{1-p}\right]$: اللوغاريتم الطبيعي لنسبة الترجيح

وبذلك يمكن كتابة معادلة الانحدار في حالة وجود عدد(k) من المتغيرات المستقلة على الصورة التالية:

$$\log_e \left(\frac{p}{1-p}\right) = \hat{B}_0 + \hat{B}_1 x_1 + \hat{B}_2 x_2 + \cdots + \hat{B}_k x_k \qquad (7)$$

$$\frac{p}{1-p} = e^{\hat{B}_0 + \hat{B}_1 x_1 + \hat{B}_2 x_2 + \cdots + \hat{B}_k x_k} \qquad (8)$$

ويتم تقدير معالم النموذج اللوجستي باستخدام طريقة الامكان الاعظم (Maximum Likelihood). ثم تبدأ مرحلة تقويم النموذج بعد تقدير معالم النموذج، (Hosmer & Lemshow, 2000) وهناك طريقتان للتحقق من ملاءمة النموذج وهي:

ا- اختبارات جودة التوفيق:

ويتم ذلك من خلال اختبار نسبة الامكان (Likelihood Test)، اختبار Hosmer-Lemeshow، جدول التصنيف (Classification table)، وتحليل منحنى ROC (Receiver Operating Characteristic)

ب- اختبار معنوية المعاملات:

وعادة ما يتم استخدام اختبار والد (Wald Test) لاختبار معنوية كل متغير مستقل على حدة وكذلك لتوضيح درجة اهمية كل متغير مستقل، حيث انه كلما زادت قيمة احصاءة والد دل ذلك على اهمية المتغير .والعكس صحيح. (King, 2002 ;Menard, 2002).

التحليل الإحصائي لبيانات الدراسة:

تم تحليل بيانات الدراسة باستخدام البرنامج الإحصائي SPSS V24 وخلص التحليل إلى النتائج التالية:

1- تم التحقق من صدق وثبات استمارة الاستقصاء من خلال اختبار الفا كرونباخ Cronbach'Alpha وقد اكدت النتائج ان قيمة الفا قد تجاوزت 0.70 مما يعنى أن الاستمارة تتسم بدرجة عالية من الصدق والثبات.

2- تم اجراء اختبار التوزيع الطبيعي (Normality Test) لجميع المتغيرات المستقلة وأشارت نتائج Kolmogorov-Smirnov أن (Sig=0.000 < 0.05) لجميع المتغيرات المستقلة مما يعنى ان المتغيرات المستقلة لا تتبع التوزيع الطبيعي.

3- النسب المئوية لمتغيرات الدراسة كما هي موضحة في الجدول رقم (3)

جدول رقم (3)

التوزيع النسبى لمتغيرات الدراسة

المتغير	التصنيف	النسب
عدد افراد الاسرة X1	(4) أفراد فاقل	47.7%
	اكثر من (4) أفراد	52.3%
متوسط الدخل الشهري للأسرة X2	4000 جنية فاقل	51%
	اكثر من 4000 جنية	49%
الحالة الوظيفية لرب الأسرة X3	لا يعمل	13.5%
	يعمل	86.5%
المستوى التعليمي لرب الأسرة X4	غير متعلم	30.5%
	متعلم	69.5%
نوعية السكن X5	إيجار	66.4%
	ملك	33.6%
وجود طلبة جامعيين بالأسرة X6	لا يوجد	62.5%
	يوجد	37.5%
وجود افراد بالأسرة لديهم مرض مزمن X7	لا يوجد	77.9%
	يوجد	22.1%
كفاية متوسط الدخل الشهري للأسرة Y	غير كاف	70.8%
	كاف	29.2%

المصدر: من إعداد الباحث

4- تم حساب مصفوفة معاملات الارتباط البسيط بين كل زوج من متغيرات الدراسة، وكانت على النحو التالي:

مصفوفة الارتباط بين متغيرات الدراسة

Variables	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Y	1	-.136	.575	.187	.051	.429	- .047	-.121
X_1	-.136	1	.066	.115	-.054	.026	.356	.166
X_2	.575	.066	1	.205	.230	.473	.188	-.070
X_3	.187	.115	.205	1	.135	.136	.008	-.027
X_4	.051	-.054	.230	.135	1	.267	-.013	-.042
X_5	.429	.026	.473	.136	.267	1	.019	-.060
X_6	- .047	.356	.188	.008	-.013	.019	1	.002
X_7	-.121	.166	-.070	-.027	-.042	-.060	.002	1

اتضح من خلال مصفوفة معاملات الارتباط عدم وجود علاقة ارتباط ذات دلالة معنوية بين المتغيرات المستقلة، من ناحية أخرى بلغت قيمة محدد مصفوفة الارتباط بين المتغيرات المستقلة = 0.529 أي لا يساوى الصفر مما يعطى مؤشر بعدم وجود مشكلة تعدد العلاقات الخطية Multicollinearity وللتأكد من ذلك تم حساب معامل تضخم التباين (Variance Inflation Factor) للمتغيرات المستقلة، وكانت على النحو المبين بالجدول رقم (4)

جدول رقم (4)
قيم معامل تضخم التباين(VIF)

Variables	Tolerance	VIF
X_1	.828	1.208
X_2	.714	1.401
X_3	.932	1.073
X_4	.902	1.109
X_5	.744	1.344
X_6	.833	1.200
X_7	.961	1.040

يتضح من الجدول انخفاض قيم معامل تضخم التباين VIF لجميع المتغيرات المستقلة اذ لم تتعدى أي قيمة من قيم VIF القيمة (10) مما يؤكد بعدم وجود مشكلة تعدد العلاقات الخطية Multicollinearity

أولاً: نتائج تطبيق أسلوب الشبكات العصبية الاصطناعية (ANN)

(1) توصيف نموذج الشبكات العصبية:

حيث أن الهدف من الدراسة هو تصنيف المشاهدات، لذلك سوف يتم الاعتماد على (Step Function) كأحد دوال التحويل، وذلك لأنها تناسب عملية التصنيف بين المشاهدات، حيث أنها تعطى نتيجتين فقط (0 ، 1) كما يتضح من خلال الصيغة التالية:

$$F(x) = \begin{cases} 0 \; when \; x \leq t \\ 1 \; when \; x > t \end{cases} \qquad (9)$$

وتصبح صيغة الدالة أو النموذج كما يلي:

$$F(x) = \frac{1}{1+e^{-s}}$$

$$S = \sum_{i=1}^{n} x_i w_i + \theta$$

حيث:

w_i:الاوزان (الأهمية النسبية للمتغيرات) ، x_i :المتغيرات المستقلة ، θ:حد التحيز(Bise)

وبذلك يأخذ النموذج الصيغة التالية:

$$F(x) = \frac{1}{1+e^{-(w_1 x_1 + w_2 x_2 + \cdots + w_7 x_7 + \theta)}} \qquad (10)$$

(2) تدريب واختبار الشبكة:

يوضح الجدول رقم (5) ملخص بعدد الحالات المستخدمة في عمليتي تدريب واختبار الشبكة

جدول رقم (5)
عدد الحالات المستخدمة في عمليتي تدريب واختبار الشبكة

Case Processing Summary			
		N	Percent
Sample	Training	258	67.2%
	Testing	126	32.8%
Valid		384	100.0%
Excluded		0	
Total		384	

يتضح من الجدول السابق استخدام (258) مشاهدة وبنسبة 67.2% من إجمالي المشاهدات في عملية تدريب الشبكة، واستخدام (126) مشاهدة وبنسبة 32,8% من أجمالي المشاهدات في اختبار الشبكة.

(3) معلومات الشبكة العصبية:

يوضح جدول رقم (6) معلومات الشبكة العصبية المستخدمة وهي مكونة من ثلاثة اجزاء:

جدول رقم (6)

معلومات الشبكة العصبية

Network Information			
Input Layer	Factors	1	عدد افراد الاسرة
		2	متوسط الدخل الشهري للأسرة
		3	الحالة الوظيفية لرب الاسرة
		4	المستوى التعليمي لرب الاسرة
		5	نوعية السكن
		6	وجود طلبة جامعيين بالأسرة
		7	وجود افراد لديهم مرض مزمن
	Number of Units[a]		14
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		7
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	كفاية الدخل
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy
a. Excluding the bias unit			

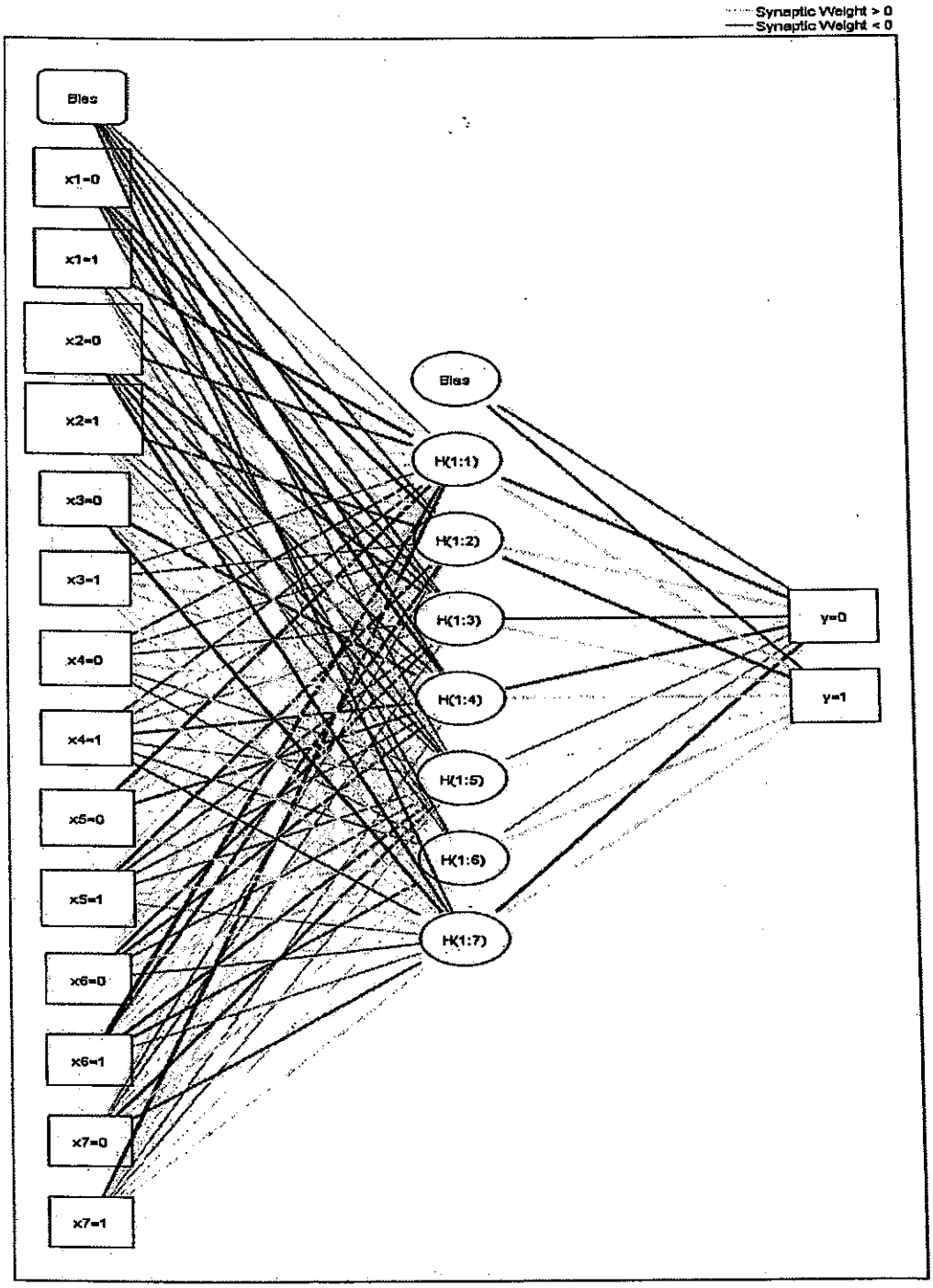
الجزء الأول: يوضح معلومات طبقة المدخلات Input Layer ممثلة في (7) متغيرات مستقلة لكل متغير مستويين (0 ، 1) وبالتالي يكون عدد الوحدات داخل طبقة المدخلات (14) وحدة.

الجزء الثاني: يوضح معلومات الطبقة الخفية Hidden Layer(s) ممثلة في طبقة واحدة، اما عدد الوحدات داخل الطبقة الخفية (7) وحدات، وان دالة التنشيط المستخدمة هي Hyperbolic tangent.

الجزء الثالث: يوضح معلومات طبقة المخرجات Output Layer ممثلة في متغير تابع واحد وهو كفاية دخل الاسرة (كافي – غير كافي) وبالتالي عدد الوحدات المعالجة بهذه الطبقة (2) وحدة، وان دالة التنشيط المستخدمة في هذه الطبقة هي Softmax وتعرف ايضا بالدالة اللوجستية Sigmoid Function

يوضح الشكل رقم (1) الشكل الهرمى لهيكل (معمارية) الشبكة المستخدمة والتي تتكون من طبقة المدخلات وتقع على اليسار مكونة من (14) وحدة بالإضافة الى وحدة التحيز (Bias) ، الطبقة الثانية هي الطبقة الخفية وتقع في الوسط وتتكون من (7) وحدات بالإضافة الى وحدة التحيز (Bias) ، أما الطبقة الأخيرة وتقع على اليمين هي طبقة المخرجات (الناتج) وتوضح أن هناك نتيجتين هما (الدخل غيركافى = 0) ، (الدخل كافي = 1)

شكل رقم (1)
الشكل الهرمي لهيكل (معمارية) الشبكة المستخدمة



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Softmax

(4) ملخص نموذج الشبكة العصبية:

يوضح جدول رقم (7) ملخص نموذج الشبكة العصبية المستخدمة

جدول رقم (7)

ملخص نموذج الشبكة العصبية

Model Summary		
Training	Cross Entropy Error	70.523
	Percent Incorrect Predictions	12.0%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.45
Testing	Cross Entropy Error	35.117
	Percent Incorrect Predictions	11.9%
Dependent Variable: كفاية الدخل		
a. Error computations are based on the testing sample.		

من الجدول السابق يتضح ما يلي:

- نسبة التصنيف الخاطئ في عينة التدريب بلغت 12%، بينما نسبة التصنيف الخاطئ في عينة الاختبار بلغت 11.9% وهي نسبة متقاربة جداً وهذا يدل على ان الشبكة تدربت بشكل جيد على تصنيف المشاهدات.
- أن قانون توقف الشبكة عن التدريب Stopping Rule Used هو عندما تصبح نسبة الخطأ ثابتة، أي عندما تتوقف نسبة الخطأ عن الزيادة.
- زمن تدريب الشبكة هو (45) ثانية.

(5) نتائج التصنيف:

يوضح الجدول رقم (8) نتائج التصنيف باستخدام الشبكة العصبية المقترحة

جدول رقم (8)

نتائج التصنيف باستخدام الشبكة العصبية

Classification				
Sample	Observed	Predicted		
		غير كافي	كافي	Percent Correct
Training	غير كافي	161	20	89.0%
	كافي	11	66	85.7%
	Overall Percent	66.7%	33.3%	88.0%
Testing	غير كافي	81	10	89.0%
	كافي	5	30	85.7%
	Overall Percent	68.3%	31.7%	88.1%

من الجدول السابق يتضح ما يلي:

- التصنيف الصحيح للدخل غير الكافي بلغ 89% في عينة التدريب، بينما بلغ 89% ايضا في عينة الاختبار.
- التصنيف الصحيح للدخل الكافي بلغ 85.7% في عينة التدريب، بينما بلغ 85.7% ايضا في عينة الاختبار.
- بلغت نسبة التصنيف الصحيح للمشاهدات باستخدام الشبكة العصبية المقترحة 88.1%، وهي تمثل نسبة جيدة جداً للتنبؤ بتصنيف المشاهدات الجديدة. وماسبق يؤكد صحة الفرض الاول للدراسة وهو إمكانية استخدام الشبكات العصبية الاصطناعية في عملية التصنيف بكفاءة عالية.

(6) الأهمية النسبية لمتغيرات الدراسة:

يوضح الجدول رقم (9) تحديد الأهمية النسبية للمتغيرات المستقلة

جدول رقم (9)

الأهمية النسبية للمتغيرات المستقلة

Independent Variable Importance		
	Importance	Normalized Importance
X1	.159	65.0%
X2	.245	100.0%
X3	.128	52.4%
X4	.132	53.8%
X5	.140	57.0%
X6	.101	41.3%
X7	.094	38.3%

من الجدول السابق يتضح ما يلي:

- اكثر المتغيرات تأثيراً في تصنيف المشاهدات باستخدام الشبكة العصبية المقترحة هو متوسط الدخل الشهري للأسرة (X2) اذ بلغت نسبة اهميته 24.5% وهذا أمر يتفق مع الواقع بشكل كبير، ثم يليه متغير عدد أفراد الأسرة (X1) بنسبة أهمية بلغت 15.9% ، ثم يليه متغير نوعية السكن (X5) بنسبة أهمية بلغت 14% ، ثم يليه متغير المستوى التعليمي لرب الأسرة (X4) بنسبة أهمية بلغت 13.2% ، ثم متغير الحالة الوظيفية لرب الاسرة (X3) بنسبة أهمية بلغت 12.8% ثم يليه متغير وجود طلبة جامعيين بالأسرة (X6) بنسبة أهمية بلغت 10.1% ، واقل المتغيرات تأثيراً في تصنيف المشاهدات هو وجود أفراد بالأسرة لديهم مرض مزمن (X7) بنسبة أهمية 9.4% .
- العمود الأخير (Normalized Importance) يمثل نسبة أهمية كل متغير مستقل إلى اكثر المتغيرات تأثيرا في تصنيف المشاهدات وهو(X2) ، فنجد على سبيل المثال أن (X1) يمثل 65% من الأهمية النسبية للمتغير (X2) 65.0% = (.159\.245) وهكذا لباقي المتغيرات.
- دالة التحويل المناسبة لبيانات الدراسة عند استخدام نموذج الشبكات العصبية هي الدالة اللوجستية لأنها تعطى نتيجتين فقط (0 ، 1) وتأخذ الصورة التالية:

$$F(x) = \frac{1}{1 + e^{-(.159x_1 + .245x_2 + .128x_3 + .132x_4 + .140x_5 + .101x_6 + .094x_7)}}$$

ثانياً: نتائج تطبيق نموذج الانحدار اللوجستي الثنائي

(1) توصيف نموذج الانحدار اللوجستي الثنائي

بما أن معادلة النموذج اللوجستي المقدرة تأخذ الشكل التالي:

$$\frac{\hat{p}(x)}{1 - \hat{p}(x)} = e^{\hat{B}_0 + \hat{B}_1 x_1 + \hat{B}_2 x_{2+\ldots+} \hat{B}_k x_k}$$

وحيث أن لدينا سبعة متغيرات مستقلة للظاهرة محل الدراسة، عليه يأخذ النموذج الشكل التالي:

$$\frac{\hat{p}(x)}{1 - \hat{p}(x)} = e^{\hat{B}_0 + \hat{B}_1 x_1 + \hat{B}_2 x_2 + \hat{B}_3 x_3 + \hat{B}_4 x_4 + \hat{B}_5 x_5 + \hat{B}_6 x_6 + \hat{B}_7 x_7} \qquad (11)$$

(2) تحليل البيانات باستخدام النموذج اللوجستي:

تم الاعتماد على نتائج تطبيق أسلوب Backward stepwise (Likelihood Ratio) وكانت النتائج كما يلي

- اشارت الاختبارات الاحصائية إلى معنوية النموذج ككل، وذلك من خلال اختبار نسبة الامكان (Likelihood Ratio Test) وكذلك اختبار (Hosmer – Lemeshow) كما يتضح من الجداول التالية:

جدول رقم (10)
نتائج اختبارات معنوية النموذج ككل

	-2LL	Chi-Square	P_value
Model	264.507	199.086	0.000

يتضح من الجدول ان قيمة (P_value= 0.000<0.05) مما يؤكد ان النموذج ككل معنوي ويمثل البيانات تمثيلاً جيداً.

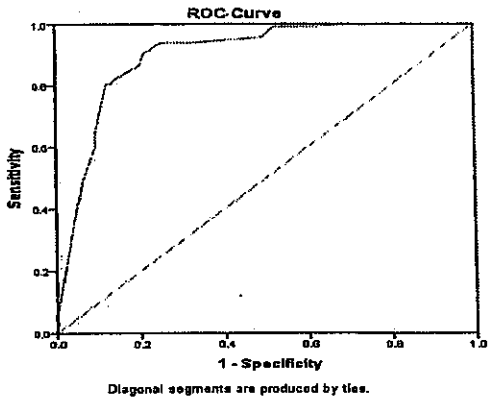
جدول رقم (11)
نتائج اختبار (Hosmer – Lemeshow)

	Chi-Square	d.f	P_value
Model	14.399	8	0.072

يتضح من الجدول أن قيمة (P_value= 0.072>0.05) مما يؤكد على عدم وجود اختلاف بين قيم المشاهدات الفعلية والقيم المقدرة مما يدل على جودة التوفيق للنموذج.

- يتضح من خلال شكل (2) والذى يمثل منحنى ROC أن النموذج يعمل على تصنيف المشاهدات أفضل مما يعمل عامل الصدفة، حيث يظهر أن المنحنى يبتعد عن قطر الصدفة والذى يحصر تحته (50%) من المساحة ليعطي مساحة اكبر مما تغطيها الصدفة، حيث ان كفاءة التقسيم الراجع للصدفة والنموذج يجب ان تزيد عن 72%

شكل رقم (2)
منحنى ROC



والجدول التالي يوضح قيمة المساحة تحت منحنى ROC للنموذج الذي تم توفيقه

جدول رقم (12)
المساحة تحت منحنى ROC

Area	S.E	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.900	0.016	0.000	0.868	0.932

يتبين من الجدول ان قيمة المساحة تحت المنحنى تساوى 0.90 عند مستوى دلالة 0.000 بفترة ثقة 95% (0.868 – 0.932) ، وهذا يعنى ان النموذج يساعد على التنبؤ بتصنيف حالات المتغير التابع اكثر مما تفعله الصدفة.

- اشارت الاختبارات الاحصائية إلى معنوية خمس متغيرات مستقلة فقط وهي تمثل أهم المتغيرات تأثيراً على كفاية دخل الأسرة في مدينة القاهرة وكانت على النحو التالي:

X1 : عدد أفراد الأسرة X2 : متوسط الدخل الشهري للأسرة
X4 : المستوى التعليمي لرب الأسرة X5 : نوعية السكن
X6 : وجود طلبة جامعيين بالأسرة

وقد اتضح معنوية هذه المتغيرات عند مستوى معنوية 5% وبالتالي تم استبعاد كل من X3، X7 كما يتضح من الجدول التالي:

جدول رقم (13)
نتائج تطبيق نموذج الانحدار اللوجستي

variables	B	S.E	Wald	d.f	Sig	Exp(B)	95% C.I .for Exp(B) Lower	Upper
X1	-1.687-	.407	17.160	1	.000	.185	.083	.411
X2	4.057	.506	64.221	1	.000	57.815	21.433	155.951
X4	-1.592-	.421	14.305	1	.000	.204	.089	.464
X5	1.401	.339	17.051	1	.000	4.060	2.088	7.894
X6	-.764-	.331	5.314	1	.021	.466	.243	.892
Constant	-2.447-	.417	34.413	1	.000	.087		

- تم اختبار معنوية معاملات المتغيرات ، واتضح معنويتها باستخدام اختبار (Wald) وكذلك قيمة (Sig) حيث أن قيمة (Sig < 0.05) للمتغيرات الخمس، وفى ضوء اختبار (Wald) نجد أن متوسط الدخل الشهري للأسرة (X2) هو اكثر المتغيرات تأثيراً واهمية في تصنيف دخل الأسرة من حيث كونه كافي أم غير كافي ، يليه المتغير الخاص بعدد أفراد الاسرة (X1) ، ثم متغير نوعية السكن (X5) ، ثم المتغير الخاص بالمستوى التعليمي لرب الأسرة (X4) ، واخيراً متغير وجود طلبة جامعيين بالأسرة (X6) ، حيث أنه كلما زادت قيمة (Wald) دل ذلك على أهمية المتغير ، وقد أشارت النتائج إلى عدم معنوية المتغير الخاص بالحالة الوظيفية لرب الأسرة (X3) ، وكذلك المتغير الخاص بوجود أفراد بالأسرة لديهم مرض مزمن (X7) وبالتالي تم حذفهم من النموذج .
- اشارت قيمة نسبة الترجيح odds Ratio (Exp(B)) إلى قيمة الدالة الأسية لمعامل الانحدار وهى تعبر عن المضاعف الذى تتغير به نسبة الترجيح (قيمة المتغير التابع) ، أي التغير من احتمال (Y =1) إلى احتمال (Y = 0) ويتم حسابها من الصيغة:

$$EXP(B) = e^{B_i} = e^{-1.687} = 0.185$$

وهذا يعنى أنه كلما زادت قيمة المتغير المستقل (X1) يقل احتمال كفاية الدخل بمقدار 18.5% تقريبا وهكذا لباقي المتغيرات ، بينما يمثل العمود الأخير بالجدول حدود الثقة لقيمة الدالة الاسية التي تم حسابها.

- يشير الجدول رقم (14) إلى أن استخدام الانحدار اللوجستي قد حقق نسبة تصنيف كلية صحيحة بلغت 85.7% وهي نسبة مرتفعة تدل على أن النموذج يمثل البيانات تمثيلاً جيداً.

جدول رقم (14)
جدول التصنيف

Observed		كفاية الدخل		Percentage Correct
		غير كافي	كافي	
كفاية الدخل	غير كافي	239	33	87.9
	كافي	22	90	80.4
Overall Percentage				85.7

- من كل ما سبق يمكننا أن نستنتج أن نموذج الانحدار اللوجستي المستخدم في تصنيف دخل الأسرة في مدينة القاهرة من حيث كونه كافي أم غير كافي يأخذ الصورة التالية:

$$\text{Log}\left(\frac{\hat{p}}{1-\hat{p}}\right) = -2.447 - 1.687\,X_1 + 4.057\,X_2 - 1.592\,X_4 + 1.401\,X_5 - 0.764\,X_6$$

النتائج والتوصيات:

أولاً: النتائج

1- من خلال التحليل باستخدام النماذج المقترحة تم التوصل إلى أن استخدام الشبكات العصبية الاصطناعية تعطى نسبة تصنيف أفضل من نموذج الانحدار اللوجستي، حيث بلغت نسبة التصنيف الصحيح باستخدام الشبكات العصبية 88.1%، بينما بلغت نسبة التصنيف الصحيح باستخدام نموذج الانحدار اللوجستي 85.7%

2- تطابقت نتائج النماذج المستخدمة من حيث أهمية المتغيرات المستقلة المؤثرة معنوياً في عملية تصنيف المشاهدات ، حيث توصلت النتائج في كلا النموذجين ان متوسط الدخل الشهري للأسرة (X2) هو أكثر المتغيرات تأثيراً في عملية التصنيف ، ثم يليه متغير عدد أفراد الأسرة (X1) ، ثم متغير نوعية السكن (X5) ، ثم متغير المستوى التعليمي لرب الأسرة (X4) ، في حين اشارت نتائج نموذج الانحدار اللوجستي إلى عدم معنوية المتغير الخاص بالحالة الوظيفية لرب الأسرة (X3) ، وكذلك المتغير الخاص بوجود أفراد بالأسرة لديهم مرض مزمن (X7) .

3- بالرغم من قدرة نماذج الشبكات العصبية على تصنيف المشاهدات بكفاءة عالية، الا انها لا تتيح لنا إمكانية اجراء الاختبارات الإحصائية المتعلقة بمعنوية العلاقة، بينما ذلك متاح في حال استخدام نموذج الانحدار اللوجستي.

4- النموذج الإحصائي المقترح استخدامه في تصنيف دخل الأسرة في مدينة القاهرة من حيث كونه كافي أم غير كافي عند استخدام نموذج الانحدار اللوجستي يأخذ الصورة التالية:

$$\text{Log}\left(\frac{\hat{p}}{1-\hat{p}}\right) = -2.447 - 1.687\,X_1 + 4.057\,X_2 - 1.592\,X_4 + 1.401\,X_5 - 0.764\,X_6$$

5- دالة التحويل التي تستخدم في تصنيف دخل الأسرة في مدينة القاهرة من حيث كونه كافي أم غير كافي عند استخدام نموذج الشبكات العصبية هي الدالة اللوجستية وتأخذ الصورة التالية:

$$F(x) = \frac{1}{1 + e^{-(.159x_1 + .245x_2 + .128x_3 + .132x_4 + .140x_5 + .101x_6 + .094x_7)}}$$

ثانياً: التوصيات

1- إذا كان الهدف من عملية تصنيف المشاهدات هو الحصول على اعلى دقة تصنيف فقط فمن الأفضل استخدام اسلوب الشبكات العصبية، اما إذا كان الهدف هو الحصول على اعلى دقة تصنيف مع شرح وتفسير لمعلمات النموذج المستخدم يمكن استخدام نماذج الانحدار والشبكات العصبية معاً. (عبد العال،2004)

2- اجراء المزيد من الدراسات حول طبيعة البيانات ومجال التطبيق للمفاضلة بين الشبكات المختلفة مع المقارنة بأساليب التحليل الأخرى.

3- الاعتماد على أسلوب الشبكات العصبية بدلا من نموذج الانحدار اللوجستي إذا كانت بعض المتغيرات المستقلة لا تتبع التوزيع الطبيعي أو لا يعرف التوزيع الذي تتبعه.

4- التوسع في استخدام الطرق والاساليب المستخدمة في عملية التصنيف في المجالات الاقتصادية والاجتماعية، وعدم حصر استخدامها على المجالات الطبية فقط كما كان في السابق.

5- اجراء المزيد من الدراسات حول دخل وأنفاق الأسر مع إدخال المزيد من المتغيرات المستقلة المؤثرة على دخل وأنفاق الأسر مثل اسعار السلع وسعر الصرف وغيرها، حيث ان عدم كفاية الدخل يؤثر بشكل مباشر على نوعية التعليم والصحة لأفراد الأسرة وجميع النواحي الحياتية الاخرى، كما يمكن ان يؤثر على حجم ومعدلات الجريمة ومستوى الأمن القومي.

المراجع

اولاً: المراجع العربية:

1- العباسي، عبد الحميد (2013)."مقدمة في الشبكات العصبية الاصطناعية وتطبيقاتها في العلوم الاجتماعية باستخدام SPSS" ، جامعة القاهرة، معهد الدراسات والبحوث الاحصائية، قسم الاحصاء الحيوي والسكاني.

2- الغصين، هلا بسام (2004). "استخدام النسب المالية للتنبؤ بتعثر الشركات"، رسالة ماجستير، غزة، الجامعة الاسلامية.

3- بدرى، مسعود؛ المطوع،احمد؛هادى، العقيل (1996)."استخدام تحليل التمايز والشبكات العصبية في التنبؤ بدرجة اعتمادية العميل المصرفي"، المجلة العربية للعلوم الإدارية، الكويت، مجلد (3)، عدد (2)، ص: 295- 315

4- جواد، عباس ناجى (2010)."المفاضلة بين طرق تقدير الدوال الاقتصادية ذات المتغيرات التابعة النوعية"،العراق، مجلة تكريت للعلوم الإدارية والاقتصادية، المجلد (6)، عدد (18).

5- دخيل، طاهر ريسان (2008)."استخدام الشبكات العصبية لأغراض التمييز "، جامعة القادسية، كلية الادارة والاقتصاد، قسم الاحصاء، المجلد (52/14) ، ص: 246 – 256

6- ريتشارد جونسون، دينوسرن (1997) - تعريب عبد الرحمن حامد عزام،"التحليل الاحصائي للمتغيرات المتعددة من الوجه التطبيقيه"، دار المريخ للنشر، الرياض.

7- عبد العال، مدحت محمد احمد (2004)."الشبكات العصبية وتطبيقات إدارة الاعمال"، القاهرة، المجلة العلمية للاقتصاد والتجارة، العدد الاول، ص :465 - 494

8- عطية، عبد القادر (2004)."الحديث في الاقتصاد القياسي بين النظرية والتطبيق"، الدار الجامعية للنشر والتوزيع، الاسكندرية، مصر.

9- على، فادية أدم (2016)."استخدام الدالة التمييزية الخطية لتحديد أهم العوامل الاقتصادية والاجتماعية التي تؤثر على كفاية دخل الأسرة بولاية شمال كردفان"، رسالة ماجستير، كلية العلوم، جامعة السودان للعلوم والتكنولوجيا.

10- عيسى، علام زكى (2000)."الشبكات العصبية – البنية الهندسية، الخوارزميات، التطبيقات"، الطبعة الأولى، شعاع للنشر والعلوم، سوريا، حلب.

11- غانم، عدنان؛ الجاعونى، فريد (2011)."استخدام تقنية الانحدار اللوجستى ثنائي الاستجابة في دراسة أهم المحددات الاقتصادية والاجتماعية لكفاية دخل الأسرة – دراسة تطبيقية على عينة عشوائية من الاسر في محافظة دمشق"، سوريا، مجلة جامعة دمشق للعلوم الاقتصادية والقانونية، المجلد (27)،العدد الأول.

12- غانم، عدنان؛ الجاعونى، فريد (2007)."التحليل الإحصائي متعدد المتغيرات (التحليل التمييزي) في توصيف وتوزيع الأسر داخل الهيكل الاقتصادي الاجتماعي في المجتمع"، سوريا، مجلة جامعة دمشق للعلوم الاقتصادية والقانونية، المجلد (23) ، العدد الثاني.

13- حمدي، محمد عبد الرحيم؛ ذنون، كرم (2009)."تشخيص مرض التدرن الرئوي (السل) باستخدام الشبكات العصبية الاصطناعية"، ورقة بحثية منشورة، مجلة الرافدين لعلوم الحاسبات والرياضيات، المجلد (6)، العدد (1).

14- نوح، أدم الطاهر (2015)."تصنيف وتحليل فئات الدخل في السودان باستخدام الدالة التمييزية مقارنة بنماذج الشبكات العصبية، رسالة دكتوراه، جامعة السودان للعلوم والتكنولوجيا، كلية الدراسات العليا.

15- الجهاز المركزي للتعبئة العامة والإحصاء (2016)."الكتاب الإحصائي السنوي".

16- الجهاز المركزي للتعبئة العامة والإحصاء (2016)."نتائج بحث الدخل والإنفاق والاستهلاك".

ثانياً: المراجع الأجنبية:

1- Bandy, H. (1994)."Thoughts on Desirable Features for a neural Network – Based Financial Trading System", NUROVST Journal, 2(3): 19-22

2- Hosmer, W. & Lemeshow, S. (2000). "Applied Logistic Regression",2nd edition New York: Johnson Wiley & Sons, Inc.

3- King, J.E. (2002). "Logistic Regression: Going beyond point-and-click", Paper Presented at the annual Meeting of the American Educational Research Association, New Orleans, LA, April.

4- King, J.E. (2003). "Running A Best-Subsets Logistic Regression: An Alternative to Stepwise Methods", Educational and Psychological Measurement, Vol. 63, No. 3, June, 392- 403.

5- Krose, B. & Smagt, P. (1996),"An introduction to neural networks", Eighth edition, The University of Amsterdam.

6- Lea, S. (1997). "Multivariate Analysis II: Manifest Variables analysis Topic 4: Logistic Regression and Discriminant Analysis ", University of EXETER Department of Psychology Available at: www.exeter. Ac.uk /_SEGL ea./multivar2/diclogi.htm1

7- Lea, S. (2004). "Application of Likelihood ratio and Logistic Regression models to Landslide susceptibility mapping using GLS". Environmental Management, Vol. 34, No. 2, 223 – 232

8- Menard, S.W (2002). "applied Logistic Regression analysis", 2nd edition Sage Publication Series Quantitative Application in the Social Sciences, No. 106, Thousand Oaks, CA: Sage.

9- Pample, F. (2000). " Logistic Regression a Primer". Sage Publication Series Quantitative Application in the Social Sciences, No.07- 132, Thousand Oaks, CA: Sage.

10- Rahman, Atta. (2009)."statistical Analysis of the Different Socioeconomic Factors Affecting the Education of N.W.F.P (PAKISTAN),"PhD, Institute of Mathematical methods in Economic (EOS), University of Technolgy Vienna, Austria, Journal of Applied Quantitative Methods, Vol (4), No (1).

11- Tang, Man-Lai. (2001)."Exact Goodness of Fit test for Binary Logistic Model". Statistica Sinica (11), The Chinese University of Hong Kong.

12- Thomas N, Anne Wangombe and Nancy Khadioli. (2010)."Alogistic Regression Model to Identify Key Determinants of Poverty Using Demographic and Health Survey Data". European Journal of Social Sciences, Vol(13), No (1).

13- Yoon, Y; Swales, G. (1993)."A Comparison of Discriminant Analysis Versus Artificial Neural Network", Journal of Operational Research ,44: 51 – 60.