استخدام بعض الطرائق الاحصائية والتصنيف الشجيري في التصنيف والتنبؤ بإفلاس الشركات مالياً

م. د. صباح منفي رضا السيد وسام سرحان ذياب
كلية الادارة والاقتصاد/ جامعة بغداد
قسم الاحصاء

1- المقدمة

تستخدم تقنيات التصنيفِ بصورة واسعة جدا في كثير من المجالات التطبيقية. ولاسيما في المجالات الاقتصادية والتحليلات المالية كما في عِلْمِ تحليلِ الزبونِ التنبؤي predictive customer analytics ، الذي يتضمن ذلك الإمتلاك، المقايضة، إحتكار وإحراز الإئتمانِ والجباياتِ. إنّ هدفَ أيّ نموذج تصنيفِ هو أنْ يُصنّفَ المشاهدات في مجموعتين أو أكثر للوصول الى التنبؤ بنتيجةً ترتبط بكُلّ مشاهدة ومثال ذلك: - متجاوب أو غير متجاوب، مقصّر أو غير مقصّر، متخبط أو غير متخبط churner الخ. نهتم في الأعمال التجارية بإمكانية التنبؤ بحسن تصرف كُلّ شركة وادئها المستقبلي، وتقنيات التصنيفِ التي تُزوّدُنا بالنماذج التنبؤيةِ للغرض لنفسه. وهناك طرق مختلفة معلميه parametric وغير معلميه non-parametric يمكن أن تُستَعملَ لحَلّ مشاكل التصنيفِ. الطرق الإحصائية المعلميه parametric في طبيعتِها تضع الفرضيات حول طبيعةِ التوزيعاتِ وتُقدر معالمَ تلك التوزيعاتِ لحَلّ المشكلةِ .

2- مشكلة التصنيفَ

إنّ هدفَ أيّ تقنية تصنيفِ هو أنْ تُقسّمَ مجموعةَ القِيَمِ المُمكنةِ للمتغير المستقل A إلى مجموعتين، A0، مجموعة القيم التي ارتبطت بتلك (المشاهدات) التي أظهرت الحدثَ (الاستجابة 0)، و A1، مجموعة القِيَم التي ارتبطت بتلك (المشاهدات) الذي أظهرت الحدثَ (الاستجابة 1). ويكون الحُكْم بالنسبة لمشاهدة جديدة للتصنيف كمحتمل لإظْهار الحدثِ 0 إذا القِيَم في المجموعةِ A0 ، يُصنّفُ ما عدا ذلك كمحتمل لإظْهار حدثِ 1 إذا القِيَم في المجموعةِ A1 . من المستحيلُ نَيْل تصنيفِ مثالي لأن مشاهدتين بالقِيَمِ المماثلةِ يُمكنُ أنْ تقدما بشكل مختلف[1]. على أية حال، يُمكنُ المَجيء (الحصول) بحُكْمَ (قاعدة) مثاليَ إذ أنّ أخطاء فقدان التصنيف misclassification تكون منخفضة بقدر الإمكان. النموذج المصمّم جيداً يَجِبُ أنْ يَعطي نسبة مئوية أعلى مِنْ النقاط العاليةِ إلى المشاهداتِ التي سَتُظهر حدثَ 1 و نسبة مئوية أعلى مِنْ النقاط المنخفضةِ إلى المشاهداتِ التي سَتظهر حدثَ 0 .

3- فروض البحث

الفرض الذي أسس عليه التصنيف للشركات هو الفرضية القائلة بأن نمط معلومات المحاسبة يختلف بالنسبة للشركات الفاشلة وغير الفاشلة.

4- أهمية البحث

تبرز أهمية الموضوع من بناء النماذج القادرة على التصنيف، والقادرة على التوقع مقدّما بفشل الشركات من قبل (سنة، سنتين و ثلاث سنوات). ويمكن عدّ هذه النماذج أنظمة معيارية، في حين أنها أنشأت على نظرية الاحتمال. إن عمليه بناء نماذج قادرة على التصنيف والتنبؤ يؤدي إلى اكتشاف حالات الفشل مما يساعد في التخطيط عند التعامل مع هكذا شركات في حال توفر النسب المالية الخاصة بها وهو موضوع هام جداً في مثل هذه الظروف الحالية لبلدنا في ظل تهافت الشركات المرتبطة بعقود إتمام الإعمار. كما تفيد البنوك والمؤسسات المالية التي تتعامل مع الشركات المقبلة على الإفلاس وعدم إمكانية البقاء. المساعدة على اتخاذ القرارات الخاصة بالإقراض والمعاملات المالية. ويساعد ذلك أيضا المدراء الماليين في الشركات باتخاذ الحيطة والحذر بما يتعلق بالوضع المالي العام لشركاتهم عن طريق التبوء باستخدام النماذج المقترحة والتي تساعدهم على تبني سلوك مالي أخر أو تعديل سياساتهم المالية بما يكفل لهم قدرتهم التنافسية في الأسواق.

5- هدف البحث

إن الهدف الأساسي للبحث هو تصنيف شركات الفاشلة وغير الفاشلة والتنبؤ بفشل تلك الشركات وذلك بالاعتماد على عدد من طرق التصنيف، ويمكن إجمال الأهداف التالية للبحث.

1- استخدام الطرق المختلفة (معلمية، تكراريه) في تصنيف الشركات.
2- التنبؤ بفشل الشركة قبل فترة زمنيه (سنه، سنتين، ثلاث سنوات).
3- المقارنة بين نماذج التصنيف المختلفة.
4- إعداد نموذج قادر على التنبؤ بفشل الشركات الأجنبية مما يساعد مبرمي العقود على تقويم تلك الشركات.

6- الجانب النظري

اولاً: الطرق المعلميه Parametric

1- الانحدار اللوجستي[10]

الانحدار اللوجستي أحد أكثر التقنياتِ المستعملةِ على نطاق واسع للتصنيفِ. الانحدار اللوجستي الثنائي Binary logistic regression أو الانحدار اللوجستي متعدد الحدود multinomial logistic regression يُمْكِنُ أَنْ يُستَعملَ عندما يكون المتغير التابعَ نوعي كـ نعم / لا متغيّراً ثنائي التفرع (dichotomous) أو يكون المتغير التابع مصنفاً (categorical) بفئات أكثر . الانحدار اللوجستي لا يقوم y (متغير تابع) بشكل مباشر. حيث يُحوّلُ المتغير التابعُ إلى متغير logit اللوغارتم الطبيعي لنسبة حدوث الحدث إلى عدم حدوثه ، حيث ln (p /1-p)) وان z = ln (p / (1-p)) هو تحويل وحدة قياس اللوغارتم logit لـ p وان :

$$z = \beta_0 + \beta X \quad \dots (1)$$

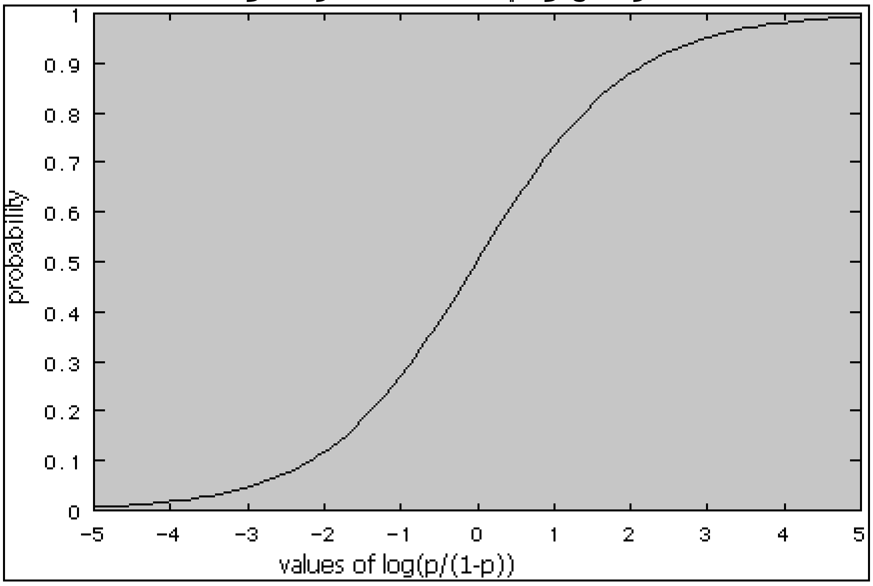
لذا فان

$$P = \frac{e^z}{1+e^z} \qquad \dots \quad (2)$$

أو

$$p = \frac{1}{1+e^{-z}}$$

ويلاحظ إن الانحدار اللوجستي يستخدم للمتغيّراتِ المـصنفة categorical variables، فان الناتج دالة مستمرة، منحناها على هيئة حرف S (شكل(1)) الذي يُمثِّلُ الاحتمال المقابل لكل صنف معيّن مِنْ المتغير التابع والتي يُمكنُ أَنْ يسبب زيَادَة أَو نقصان بمنحنى الدالة .



شكل (1) الدالة اللوجستية

تستخدم طريقه الإمكان الأعظم (MLE) لتَقدير معاملاتِ النموذج إذ توجد تلك التقـديرات لجعل دالة الإمكان أعظم ما يمكن، وفي حالة (فشل ونجاح الشركات) فان المتغير y_i والذي يمثـل مجموع حالات الفشل . ان داله الإمكان لـ n من المشاهدات هي :

$$L = \prod_{i=1}^{n} \left(\frac{e^{x'_i\beta}}{1+e^{x'_i\beta}} \right)^{\sum_{i=1}^{n} Y_i} \left(1 - \frac{e^{x'_i\beta}}{1+e^{x'_i\beta}} \right)^{n-\sum_{i=1}^{n} Y_i}$$

وداله الإمكان الأعظم اللوغارتمية هي :

$$l = \sum_{i=1}^{n} \left[Y_i \log\left(\frac{e^{x'_i\beta}}{1+e^{x'_i\beta}} \right) + (1-Y\ ) \log\left(1 - \frac{e^{x'_i\beta}}{1+e^{x'_i\beta}} \right) \right] \quad \ldots \quad (3)$$

وباستخدام المشتقه الأولى بالنسبة للمعاملات β_k ومساواتها للصفر وبحل k+1 من المعادلات الناتجة والتي لا يمكن حلها بالتحليل ولذا نستخدم الطرق العددية مثـل طريقـة نيـوتن رافسن للحصول على مقدرات β_k . الانحدار اللوجستي لهُ عِدّة تشابهات بالإنحـدار الخطـي. فمعاملات Logit مماثلة إلى معاملاتِ البيتا في معادلةِ الإنحدار الخطي، كما إن معاملات logit

القياسية مماثلة إلى البيتا الموزونة، وإحصاءة R^2 لبسيدو pseudo R^2 statistic تـستخدم هنا لتَلخيص قوّةِ العلاقةِ حيث:-

$$R^2 = 1 - [LL(\beta_0,\beta)/LL(\beta_0)] \quad ... \quad (4)$$

و تـستخدم الصيغة أعلاه في حالة المتغيرات النوعية. أما المعالم β_k فيمكن اختبارها على انفـراد وفقـا للصيغة (Wald test)

حيث

$$\frac{\hat{\beta}_j - \beta_{g0}}{s\hat{e}(\hat{\beta})} \sim N(0,1) \quad ... \quad (5)$$

وتقدر $s\hat{e}(\hat{\beta})$ عند تحديد معكوس مصفوفة المعلومات. الانحدار اللوجستي لا يَفتـرضُ خطيـةَ العلاقةِ بين المتغير المستقل والتابع، ولا يَتطلّبُ توزّيعَ المتغيّراتَ طبيعيا، لا يَفترضُ تجانس التباين homoscedasticity، ولهُ متطلبات أقل صرامة مِنْ الإنحدار الخطي عُموماً.

⌖ فرضيات النموذج

- الأخطاء مستقلة
- لا تعدديه خطيه Multicollinearity
- المتغير التابع تصنيفي Dependent is categorical
- إدراج كُلّ المتغيّرات ذات العلاقة وإستثناء كُلّ المتغيّرات ليست ذات علاقة.

الانحدار اللوجستي يُمْكِنُ أَنْ يُعالجَ اللاخطيةَ من دون خسران ميزاتِ النمـوذج الخطـي. و نموذج الانحدار اللوجستي حصين robust، وأقل صرامة في فرضياتِه بالمقارنة مع الإنحـدار الخطيُ وسهلَ الفَهْم والتّطبيق. وعليه، فهو إلى حد بعيد التقنية الأكثـر اسـتخداماً فـي مجـال التصنيفِ واستُعملت كثيرا في مجال عِلْم التحليل التنبؤي predictive analytics .[4]

وبعد الحصول على قيمه P في المعادلة (2) يمكن حينئذ بناء نموذج للتصنيف فإذا كانـت P>C فان المفردة تصنف على أنّها 0 وخلاف ذلك تصنف على أنّها 1 .

2- التحليل التمييزي Discriminant Analysis

يُستَعملُ التحليل التمييزي[4] Discriminant analysis لتَـصنيف المـشاهدات إلـى اثنتين أو أكثر من المجموعات المتعارضة بالاعتماد على المعلومات المقدمة من قِبل مجموعـة المتغيرات التنبؤية predictors مقابل المتغير المستقل في الانحدار، عندما لا يوجد ترتيـب طبيعيَ يقدم للفصل بين المجموعاتِ بصورة كافيه. وتتلخص طريقـه فـشر التـي قـدمت سـنه 1936[Tim] في تحويل المشاهدات المتعددة المتغيرات X الى مشاهدات وحيدة المتغير L والتي نحصل عليها من المجتمعين A_1, A_2 بقدر المستطاع، وذلك باستخدام توليفه خطية من X وكمـا موضح بالعلاقة :

$$L = b_1x_1 + b_2x_2 + ... + b_nx_n + c \quad ... \quad (6)$$

للحصول على L حيث أن b_i معاملات التصنيفdiscriminant coefficient، و x_i متغيرات مستقلة و c ثابت و إنّ المعاملاتَ (b_i)تَحْسبُ بحيث تجعل المسافة بين متوسـطات المتغيـر التابع أعظم ما يمكن . فإذا كانت μ_{1L} تشير إلى متوسط قيم L التي حصلنا عليها من قيم X التي تنتمي للمجتمع A_1 وكانت μ_{2L} تمثل متوسط قيم L التي نحصل عليها من قيم X التـي تنتمـي للمجتمع A_2. ونلاحظ ان القيمة المتوقعة لمشاهدة متعددة المتغيرات من المجتمع A_1 هي:

$$\mu_1 = E(X/A_1) \quad ... \quad (7)$$

كما ان القيمة المتوقعة لمشاهدة متعددة المتغيرات من A_2 هي:

$$\mu_2=E(X/A_2) \qquad \dots (8)$$

وبافتراض إن المجتمعين لهما نفس مصفوفة التباين والتباين المشترك حيث

$$\sum = E(X-\mu_i)(X-\mu_i)' \qquad i=1,2$$

ونجد ان متوسط L هو احد المتوسطين التاليين

$$\mu_{1L}=E(L/A_1) \qquad \dots (9)$$

$$\mu_{2L}=E(L/A_2) \qquad \dots (10)$$

وذلك يتوقف على المجتمع الذي ينتمي إليه L ، وان فشر اختار التوليفة الخطية التي تعظم مربـع مسافة مهالانوبيس Mahalanobis distance (عدد الانحرافـات المعياريـة عـن المركـز centroid) بين المفردات case والمركز لكلّ مجموعـة مـن المتغيـر التـابع. إنّ مركـز المجموعة centered هو القيمةُ المتوسطةُ لدرجات التمييز للتـصانيف category المعطـاة للمتغير المعتمد (التابع). تُصنّفُ عَائدية (تبعية) المفردة إلى المجموعـةِ للتـي لهـا مـسافةِ Mahalanobis الأصغر ، حيث يتم الحصول على أفضل توليفة :

$$\max_{\ell} \frac{(\ell\delta)^2}{\ell\sum\ell} = \delta' \sum\delta \qquad \dots (11)$$

حيث إن $\sum$ أكيدة الايجابية وان $\delta = (\mu_1 - \mu_2)$ ، وان

$$L_\circ = (\mu_1 - \mu_2)' \sum^{-1} x_\circ \qquad \dots (12)$$

هي قيمة دالة التمييز عند مشاهدة جديدة $x_\circ$ ، وان

$$m = \frac{1}{2}(\mu_{1L} + \mu_{2L}) = \frac{1}{2}(\mu_1 - \mu_2)' \sum^{-1} (\mu_1 + \mu_2) \dots (13)$$

وهي نقطة المنتصف بين متوسطي المجتمعين وحيدي المتغير ، وعليه يمكنا إن نلاحظ

$$E(L_\circ / A_1) - m \geq 0$$

$$E(L_\circ / A_2) - m < 0 \qquad \dots (14)$$

وهذا يعني أنّه لو كانت $x_\circ$ تنتمي إلى A_1 فمن المؤمل أن تكون $L_\circ$ اكبر من نقطة المنتصف، أما إذا كانت $x_\circ$ تنتمي إلى A_2 فمن المؤمل أن تكون $L_\circ$ أصغر من نقطة المنتصف لـذا فـان قاعـدة التصنيف تصنف $x_\circ$ في A_1 إذا حدث

$$L_\circ = (\mu_1 - \mu_2)' \sum^{-1} x_\circ \geq m \qquad \dots (15)$$

وبخلاف ذلك فان $x_\circ$ يمكن تصنيفها ضمن A_2 أما في حاله التساوي فان القيمة تصنف في إحدى المجموعتين عشوائيا، وباستخدام المقدرات غير المتحيزة للأوساط وللتباين نجد

$$\hat{m} = \frac{1}{2}(\bar{l}_1 + \bar{l}_2) = \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' S^{-1}_{pooled} (\bar{x}_1 + \bar{x}_2) \qquad \dots (16)$$

وعليه فان التحليل التمييزي Discriminate يُحاولُ إيجاد قاعدة التي يَفْصلُ بها العناقيدَ إلى أقصى حدّ مدى مُمكن.

⊕ فرضيات النموذج

- خطية
- قابلية جمع
- قياس صحيح مِنْ المشاهداتِ
- تساوي احتمالات نقاطِ العينات التي تَعُودُ إلى المجموعاتِ المُعَرفةِ
- تُوزّعُ الأخطاء بشكل عشوائي
- لاتعدديه خطيه multicollinearity .

العائق المهم فِيْ التحليل التمييزي هو اعتماده على توزيع مساوٍ نسبياً مِنْ عضويةِ المجموعةِ. وإذا كانت مجموعةِ واحدة ضمن المجتمع أكبرُ جوهرياً مِنْ المجموعةِ الأخرى، كما هو الحال عادةً في الحياةِ العملية، التحليل التمييزي قَدْ يُصنّفُ كُلّ المشاهدات في مجموعةِ واحدة فقط. يَجِبُ أَنْ تَختَارَ عيّنة متساوية من النوع (جيد سيئ) لبناء نموذج التحليل التمييزي. إن التقييد الهامّ الآخر للتحليل التمييزي هو عدم استطاعته مُعَالَجَة المتغير المستقل الترتيبي[14]. والتحليل التمييزي أكثرُ تصلباً مِنْ الإنحدار اللوجستي في فرضياتِه. فبالمقارنة مع الإنحدار الخطي العادي نجد أن التحليل التمييزي لا يمتلك معاملاتُ وحيدةُ. حيث إن كُلّ مِنْ المعاملاتِ تَعتمدُ على بعضها البعض في التقدير ولذا ليس هناك طريقة لتَحديد القيمة المطلقةِ لأيّ معامل. وتقدر معاملات النموذج بطريقه المربعات الصغرى المرجحة أو بطريقة الإمكان الأعظم السالفة الذكر أو بطريقة χ^2 minimum والصيغة العامة لتقدير المربعات الصغرى العامة هي :

$$B = \left(X'X\right)^{-1} X'Y \qquad \dots \quad (17)$$

وقد تم استخدام أسلوب الـ stepwise في تحديد نموذج التميز وذلك بسبب أهميته وكفاءته بتحديد المتغيرات المعنوية الداخلة بالعلاقة. إذ يدخل متغيرٌ واحدٌ في كل خطوة وأن أسلوب المتسلسل الأمامي يبدء بإدخال المتغير الأكثر تمييزا ومن ثم الزوج الذي يحقق أعظم تمييزاً من بين الأزواج الممكنة من المتغيرات وهكذا وصولا إلى الدالة المميزة التي تشتمل على العدد الأكفأ من المتغيرات في النموذج. ويمكن حساب معدل الخطأ الظاهر للتصنيف باستخدام ما يسمى بمصفوفة الخلط (confusion matrix) [7] والتي تبين الانتماء الفعلي مقابل الانتماء المتنبأ به بالنسبة لكلّ مجموعه، فإذا كان لدينا n_1 من مشاهدات المجتمع A_1 ولدينا n_2 من مشاهدات المجتمع الثاني A_2 فان مصفوفة الخلط تأخذ الشكل التالي:

الانتماء المتنبأ به

الانتماء الفعلي		A_1	A_2	
	A_1	N_{1C}	$N_{1W}=N_1-N_{1c}$	N_1
	A_2	$N_{2W}=N_1-N_{2c}$	N_{2C}	N_2

حيث:-

N_{1C}: عدد

N_{1w} : عد

N_{2w} : عدد مفردات A2 التي صنفت خطا على إنها من A1

N_{2c} : عدد مفردات A2 التي صنفت صوابا على إنها من A2

وعليه فان معدل الخطأ الظاهر هو:

$$APER = \frac{N_{1W} + N_{2W}}{N_1 + N_2} \qquad \dots \quad (18)$$

والذي يمثل نسبة مشاهدات العينة التدريبية التي صنفت خطأً إن الصيغة أعلاه تكون متحيزة في حالة العينات الصغيرة وتجدر الإشارة هنا إلى استخدام طريقة الاستبعاد للاشنبرش (Lanchenbruchs holdout procedure) إذ يتم استبعاد قيمه من المجموعة الأولى ويتم إيجاد الدالة من المشاهدات المتبقية وعددها n_1-1+n_2 ويتم تصنيف المشاهدة المستبعدة ويكرر هذا الإجراء (استبعاد وتصنيف) حتى يتم تصنيف جميع مشاهدات المجموعة الأولى، وبافتراض إن $n_{1M}^{(H)}$ تشير إلى عدد المشاهدات المستبعدة (H) من المجموعة الأولى M1 بتكرار العملية على المجوعة الثانية M2 حتى يمكن ان نحصل على عدد $n_{2M}^{(H)}$ المشاهدات المصنفة خطا من المجموعة الثانية وبتحديد الاحتمال الشرطي للتصنيفين لاخطأ التصنيف

حيث

$$\bar{E}(AER)=\frac{n_{1M}^{(H)}+n_{2M}^{(H)}}{n_1+n_2} \qquad \dots \quad (19)$$

وتدعى الطريقة أيضا بطريقة jackknifing. وتقاس دقه التمييز للنموذج بشكل عام عن طريق Ψ (wilks)[9] وهي نسبة قيمة مصفوفة التباين المشترك لداخل المجموعات $|M|$ لقيمة مصفوفة التباين والتباين المشترك $|\Sigma|$ وقيمته محصورة بين (0، 1)

حيث

$$\Psi=\frac{|M|}{|\Sigma|} \qquad \dots \quad (20)$$

وانه كلما اقتربت قيمة Ψ من 1 كلما كانت متوسطات المجاميع متساوية مما يعكس فشل الدالة في التصنيف أما إذا اقتربت النسبة من الصفر كان ذلك مؤشرا جيدا لجودة الدالة للتصنيف ، وأن كثيراً من البرمج الجاهزة التطبيقية الاحصائية تعتمد هذا المقياس كما هو الحال في برنامج SAS و SPSS 13. وقد قدم Rao اختبار F وهو :

$$F=\frac{1-\sqrt[5]{\Psi}}{\sqrt[5]{\Psi}}\,\frac{df_1}{df_2} \qquad \dots \quad (21)$$

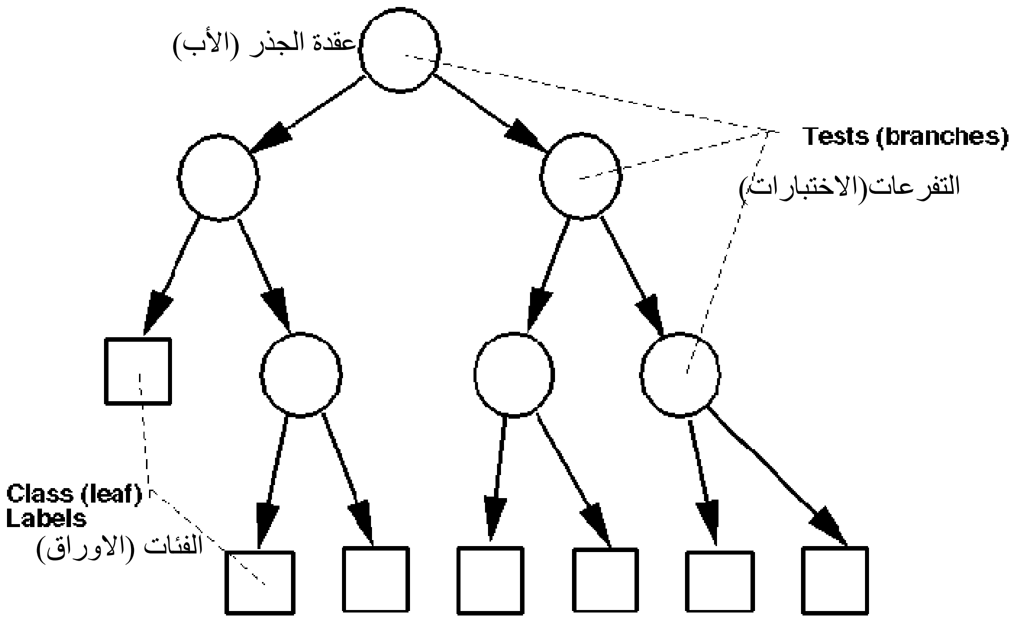
وان $df_1=ms-2\lambda$ و $df_2=p(k-1)$ حيث p عدد المتغيرات، k عدد المجاميع، وان

$$\lambda=\frac{p(k-1)-2}{4} \qquad ,s=\sqrt{\frac{p^2(1-k)^2-4}{(1-k)+p^2-5}} \qquad ,m=N-1-\frac{1}{2}(N-p)$$

وكلما كانت قيمة F كبيرة دل ذلك على معنوية النموذج .

ثانياً : طرق التَقسيم (العزل) التكرارية[8]

ومن هذه الطرق طريقة التصنيف الشجيري وتقسم بموجب هذه الطريقة مجموعه البيانات datasets إلى اثنين أو أكثر من المجاميع الفرعية لتَحسين تصنيف المتغيّر الهدف (شـكل(2)). وكُلّ مجموعة فرعية تُفْحَصُ للرُؤية فيما إذا كان بالإمكان أَنْ يَكُونَ هناك تقسّيماً آخرَ إلى مجاميع فرعيةِ داخليه حتى تحقق حالة شرطية conditional ties مُحدّدة (قواعدِ إيقاف) تتعلّقُ بحجـم العيّنةِ، والمعنوية،.... الخ. و الفكرةَ الأساس هـي أَنْ تَقـسمَ مجموعـةَ البيانـاتِ التنبؤيـة predictor إلى مجاميع جزئية subsets، كلّ منها أكثرُ تجانساً في سلوكِ أعضائه مِنْ المجموعةِ الأصليةِ .



الشكل(2) يمثل شكل التصنيف الشجري

وهكذا شجرة التصنيفِ تعتبر قاعدةُ تجريبيةُ لتَصنيف المتغير التابع مِـنْ قِـيَمِ المتغيّـراتِ التنبؤية predictor .[8] ، هناك عِدّة طرق لـشجرةِ تـصنيفِ ظَهـرت خـلال العقـد الأخيـر شـملت (C&RT) او (CART) (Classification And Regression Tree) وأيـضا QUEST, C5.0, CHAID NewID, Cal5 [Michie et al (1994)] وبصورة عامـه إن أكثر طرق أشجار التصنيفِ لها الخطوات التالية :

- دَمْج Merging - تجميع أصناف المتغير التنبؤي predictor غير المعنوية بأصناف المتغير التنبؤي predictor المعنوية نسبة إلى متغيّر الهدفَ
- إنْشِقاق Splitting - تَقسيم العقدة باختيار المتغيّر التنبؤي .
- إيقاف stopping– قواعد (حكم) Rules لتَقْرير المدى الذي يُمْكِنُ أَنْ تُقسَّمَ العُقَدِ فيه.
- تقليم Pruning - يُزيلُ الفروعَ التي لا تُضيفُ إلى القدرةِ التنبؤيةِ للنموذج.

إنّ القوّةَ العظمَى لمنهج شجرةِ التصنيفِ شفافيتُها وسهولتُها التطبيقية. وأشجار التصنيفِ يُمْكِنُ أَنْ تُمثِّلَ تقريباً بمجموعة من الاختبارات المنطقية. وتعد أشجار التصنيف ممتازة أيضاً في تَمييز جيوبِ pockets تركيزِ الحدثِ العالي والمنخفض.

فاذا كانت لدينا k من الفئات (C1,C2,. . . ,Ck) ، وعينه تجريبية من البيانات T، Training set، يمكن ملاحظة التالي:

- اذا كانت T تحتوي على واحدة أو أكثر من المشاهدات المنتمية إلى فئة واحدة Cj ستكون الشجرة ورقه Leaf مخصصه للفئة Cj .
- إذا لم تحوي T على مشاهدات لتلك الفئات فلا توجد هناك شجره لهذه البيانات .

اذا احتوت T على خليط من المشاهدات لتلك الفئات، سيكون هناك اختبار Test مبنيا على الصفات المفردة لتلك المشاهدات التي تعطي واحدة أو أكثر من النتائج المنفصلة مثنى مثنى (O1,O2,…,On) والمجموعة T ستقسم إلى مجاميع فرعية T1,T2,…,Tn حيث Ti تحوي جميع المشاهدات التي لها النتائج Oi من لاختبار المختار، وتكرر هذه العملية على جميع المجموعات الفرعية من بيانات الاختبار training set .

إنّ ضعفَ أشجار التصنيف يتمثل في حاجاتها الكثيرة للبياناتَ data hungry ويُمْكِنُ أَنْ تستغرق كثيراً مِنْ الوقتِ لبناء النماذج. ويلاحظ ان بَعْض تقنياتِ شجرةِ التصنيفِ مثل (CART, QUEST) يُمْكِنُ أَنْ تَبْني أشجاراً ثنائيةَ (شكل 2) فقط، وبمعنى آخر يُمْكِنها أَنْ تُقسّم مجموعة البيانات dataset إلى عقدتين في كلّ مّرة. والمصنفات الشجرية السالفة الذكر تختلف في تحديد الكيفية التي يتم بها تقسيم عينة التجريب training sample إلى مجاميع جزئيه وعليه فهي مختلفة في الأشجار الفرعية sub tree مما يؤدي إلى اختلاف في معايير تقويم جودة الانشقاق إلى مجاميع جزئية. وتوزع المشاهدات خلال الشجرة إلى مجاميع من القيم المقبولة acceptance ومجاميع من القيم المرفوضة rejection عند العقد الطرفية عن طريق الاختبار.

معايير الانشقاق

يتم اختيار الخواص (المتغيرات) التي تحدث أعظم ربحاً بالمعلومات بغية التصنيف الجيد وفقا لهذه الاستراتيجية وبافتراض وجود الفئتين P, N وليكن S_i الذي يحوي على عناصر p_i من الفئة P والعناصر n_i من الفئة N . ان كمية المعلومات التي نحتاجها لتقرير فيما إذا كانت خواص المفردات في المجوعة S تنتمي إلى الفئة P أو الفئة N تكون معرفة بالشكل التالي:

$$I(p,n) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{n}{p+n} \quad \ldots \quad (22)$$

وبافتراض استخدام الخاصية A ، المجموعة S ستقسم إلى V من المجاميع الجزئية $\{S_1, S_2, \ldots, S_V\}$ وإذا احتوت S_i على p_i من المفردات التابعة للفئة P وعلى n_i , من المفردات التابعة للفئة N فأنه يمكن تعريف قيمة الطاقة (Entropy) وفقا للصيغة :

$$E(A) = \sum_{i=1}^{V}\frac{p_i + n_i}{P+n}\ I(p_i, n_i) \quad \ldots \quad (23)$$

والمعلومات المحصولة عليها بواسطة التفرع وفقا للخاصية A ويمكن الحصول عليها وفقا للصيغة التالية:

$$Gain(A) = I(p,n) - E(A) \quad \ldots \quad (24)$$

وتجدر الإشارة إلى وجود دوال مشابهة أخرى لتحديد مقياسا للانشقاق وفقا للخواص المدروسة إذ يعد مؤشر جيني Gini Index الذي استحدثته شركة IBM العريقة في مجال التقنية والمعلومات مؤشراً جيداً، فاذا احتوت مجموعة البيانات T على n من الفئات فانّه يعّرف وفقًا للصيغة التالية :

$$Gini(T) = 1 - \sum_{j=1}^{n} p_j^2 \quad \ldots \quad (25)$$

حيث p_j هو تكرار نسبي للفئة j في المجموعة T .

فإذا كانت مجموعة البيانات T مقسمة إلى مجوعتين T_1 ، T_2 بحجم N_1, N_2 على التوالي فان مؤشر Gini يقسم البيانات وفقا للصيغة التالية :

$$Gini_{siplet}(T) = \frac{N_1}{N}Gini(T_1) + \frac{N_2}{N}Gini(T_2) \quad \ldots \quad (26)$$

والخاصية التي تبدي اقل قيمة لـ $Gini_{siplet}(T)$ تختار لتقسيم العقدة (ويحسب لكل الخواص المدروسة للحصول على قيم نقطة الانشقاق الممكنة لكل خاصية).

إن عملية استخراج قواعد التصنيف من الأشجار يتم باستخدام قواعد IF THEN إذ أن كل قاعدة تنشأ لكل مسار من الجذر إلى الورقة وعلية فإن كل ورقة طرفية ستحوي التنبؤ.

وباعتبار T_ℓ مجموعة للأشجار الفرعية sub tree، وان ϕ هو عدد هذه المجموعة وان العينة الاختبارية tasting sample مستخدمة لانتقاء المصنف ϕ_{bset} من Ti ويتم هذا بتصغير عدد أخطاء التصنيف $\hat{L}_{\ell,m}(\phi)$ والذي يرتبط بالعينة الاختبارية باختيار $\phi \in T_\ell$ وان

$$\phi_{test} \min_{\phi} \hat{L}_{\ell,m}(\phi) = \frac{1}{m} \sum_{i=\ell+1}^{\ell+m} I_{\{\phi(X_i)\neq y\}} \qquad \ldots \quad (27)$$

إذ أن $\phi(x_i)$ هي شجرة ثنائيه التي تخطط maps المتغيـرات التوضـيحية explanatory وأن $I_{\{\phi(X_I)\neq y\}}$ دالة المؤشر تأخذ القيمة 1 في حالة عدم التـصنيف والقيمـة 0 فـي بقيـة الحالات. ويلاحظ ان أشجار تصنيفٍ لا تُخصّصُ درجة "score" لكل مشاهدة بل يـتم تخـصيص عقدة مع أمثالها من المشاهدات ويُمْكِنُ أَنْ يَكُونَ هذا أحيانـاً ضـعفاً - ولاسـيما أن الإداريـين مُتعودون عَلى مفهومِ الدرجة "score" لاتخاذ القراراتِ. أن تقنيات شجرةِ التصنيفِ يُمْكِنُ أَنْ تحـدث مـا يـدعى بفـيض النمـوذج overfit، مـع ذلـك عبـر التـصديق cross-validation "التحقق" يُمْكِنُ أَنْ تعالجَ هذه القضيةِ. وقـد اسـتخدم برنـامج SPSS13 وبرنامج STATISTICA 6 لإنجاز شجرة التصنيف.

7- الجانب التطبيقي

لقد كان التنبؤ بفشل الشركات خلال تصنيف الحالات والتعميم المعروف إلى الحالات الأخر موضوع دراسة لـ 30 سنة ماضيه تقريبا. ويعد التنبؤ الدقيق لفشل الشركات على درجة من الأهمية بالنسبة للمستثمرين والدائنين والمدققين. ويمكن أن يكون ذا فائدة لحملة الأسهم والدائنين وللحكومات على حد سواء لتجنب الخسائر الجسيمة الناجمة عن الإفلاس المفاجىء. لذلك، فباستعمال الأدوات التحليلية والمعلومات المستخلصة من التقارير المالية للشركات نستطيع تقييم بل التنبؤ بمنزلتها المالية المستقبلية. ومع أن فكرة كون الشركة عاجزةً عن دفع ديونها متلازمة مع زوالها، فقبل حدوث ذلك حقيقة، لابد أنْ تكون الشركة قد مرت بأزمة طويلة الأمد ذات مراحل متعددة. يصنفها العديد مِنْ الباحثين في الشؤون المالية إلى مرحلتين رئيسيتين تأخذان في الحسبان اثنين من مفاهيم الفشل، هما: مالي واقتصادي، يبدأ الفشل الاقتصادي عندما تكون ربحية رأس المال المستثمّر أدنى من كاف الفرصة، فينخفض تبعا لذلك إيراد استثمار المالك فيصبح أقل من الفرص الأخرى البديلة بوجود نفس الخطر. وعند تفاقم الفشل الاقتصادي ويبدأ استقراره في الشركة، تبدأ الإيرادات بالانخفاض لتصبح أقل من الحساب، تظهر أول النتائج السلبية.[11] وإذا لم يُعالج التدهور الحاصل إبان عملية الفشل الاقتصادي، فسيقود الشركة إلى الإفلاس التقني. وهذه هي أول مرحلة مما يسمى بالفشل المالي. وفي هذه الحالة لا تملك الشركة سيوله كافية للدفع مع ازدياد تلك العوامل. سيتم الوصول إلى نقطة انقطاع لهذه العملية المدمرة لتصبح الشركة ليس فقط غير قادرة على دفع مستحقاتها بل تصبح أيضا في حالة صافي الدخل السلبي. ويعني ذلك ان التزاماتها الجارية اكبر من قيمة أصولها، لذلك تسرع هذه العملية من زوال الشركة. وأن عملية دراسة فشل الشركات يجب أن تكون من خلال البحث عن مسبباتها والتي يمكن أن نحللها من خلال علاماتها الواضحة، إن فكرة الفشل، وبدقة أكثر، العجز عن إيفاء الديون، قد بقيت مرتبطة بتقنية حسابات النسب. ويعتقد أن النسب تسوء إذا ما اقتربت الشركة أكثر من الأزمة، بهذا الشكل يمكن قياس التدهور الذي تعاني منه الشركة. وأن تحليل النسب المالية الذي يجمع كل تلك المعلومات، يُعد التقنية الأكثر شيوعا، والفائدة الكبيرة وراء المقارنة بين شركات مختلفة (في قطاع الصناعة) وهناك عقبتان رئيسيتان تتعلقان بالنسب المالية تؤثران في الاستفادة من تلك المقارنة: تكوين وأداء النسب المالية. وهناك عقبة ثالثة تُضاف إلى ما تقدم ألا وهي نفس قيمة النسبة لشركتين من قطاعين مختلفين قد لا تمثلان الموقع نفسه. ويجب أن تتجانس المعلومات المالية الملتقطة بالنسبة نفسها، لكي يتم استعمالها في وصف الشركة والتنبؤ بفشلها.

وعلى الرغم من إهمال استعمال الأساليب الإحصائية. لما يقارب من نصف قرن من قبل المحللين، فان استعمال تقنيات الإحصاء في الوقت الحاضر قد أصبح أداة مساندة شائعة لأنها تعطي موضوعية للتحليل. أن عملية الحصول على العينة و انتقاء النسب المالية مع النماذج التي تتلاءم مع موضوع الدراسة كان محورا لعديد من البحوث لفترة طويلة وان عملية اختيار النسب التسعة المستخدمة كان نتيجة التوصية التي تم التوصل إليها في بحث الدكتور لزيرجا وتأثيرها على الفشل. إن عينة الدراسة المتكونة من 120 شركة نصفها صنفت كشركات فاشلة، والنصف الأخر صنف كشركات متعافية. وقد تم اختيار عينة وذلك باستخدام قائمة من الشركات الفاشلة المختارة سابقا وربطها مع شركات معافاة وتكون هذه الشركات المعافاة والفاشلة والتي تم اختيارها بعملية الربط من الحجم نفسه و بالقطاع الصناعي. وتُبرر عملية الربط هذه لتفادي أيّ تأثير تشويه محتمل متعلق بالحجم وقطاع الصناعة، والمتغيرات التي حصلنا عليها وهي 9 من النسب التي تم استخدامها في بناء النماذج هي:

1. x_1= موجودات متداولة/ خصوم جارية

X1. CURRENT ASSETS / CURRENT LIABILITIES

2. x_2= موجودات متداولة/إجمالي الموجودات

X2. CURRENT ASSETS / TOTAL ASSETS

3. x_3= نتائج نهائية/إجمالي الموجودات

X3. NET RESULT / TOTAL ASSETS

4. x_4=دخل قبل الفائدة والضرائب / اتفاقات مالية

X4. EARNINGS BEFORE INTEREST AND TAXES / FINANCIAL CHARGES

5. x_5= الأموال الخاصة / إجمالي ديون.

X5. OWN FUNDS / TOTAL DEBT

6. x_6= مبيعات / أموال خاصة

X6. SALES / OWN FUNDS

7. x_7= أسهم / مبيعات

X7. STOCKS / SALES

8. x_8= مدينون / مبيعات

X8. DEBTORS / SALES

9. x_9=السيولة النقدية الفعّالة /إجمالي الموجودات

X9. OPERATIVE CASH FLOW / TOTAL ASSETS

نتائج التحليلات الإحصائية

1- التحليل المميز Dscriminant Analysis (DA)

باستعمال برنامج SPSS 13 طبقت الدالة المميزة على البيانات. وفيما يلي دوال المصنفات بعينة من 120 شركة ، لكل سنه من السنوات الثلاثة المتتالية قبل الفشل ، باستعمال (DA) . إذ تم الحصول على الدوال التالية بعد تنفيذ أسلوب الـ STEPWISE
لثلاث سنوات قبل الفشل والنموذج المقترح هو:-

$$L = -2.OO81 + 2.246X_4 + 2.529X_5$$

X4= دخل قبل الفائدة والضرائب/ النفقات المالية ، X5 = الأموال الخاصة / إجمالي الديون
L: درجة الفشل L: Score of Failure ، واذا كانت 0<L تصنف الشركة على إنها فاشلة ، وبالعكس فإنها ليست فاشلة ، ويلاحظ عدم معنوية عدد من المتغيرات حيث كان تأثيرها ضعيف على الدالة. وان قيمة الحد الفاصل m بعد التعويض عن قيمة الأوساط المستخرجة لكلا المتغيرين x_4, x_5 نحسب من العلاقة ونجد قيمتها مساوية إلى 2.0081 . والجدول التالي يوضح المتغيرات الداخلة ومعنوية النموذج بالاعتماد على Ψ وعلى F وكالاتي:

Variables Entered/Removed(a,b,c)

Step	Entered	Wilks' Lambda							
		Statistic	df1	df2	df3	Exact F			
						Statistic	df1	Df2	Sig.
1	x5	0.754	1	1	118	38.593	1	118	.000
2	x4	0.581	2	1	118	26.010	2	117	.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.
a Maximum number of steps is 18.
b Minimum partial F to enter is 3.84.
c Maximum partial F to remove is 2.71.

حيث يلاحظ أهمية كل من المتغيرين في المعادلة وحسب طريقة stepwise.

Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	x	1.000	38.593	
2	x	.982	27.093	.853

وان نسبة الخطأ في التصنيف بينٌ إذ أن نسبة المفردات المصنفة بصورة صحيحة تبين دقة التصنيف للنموذج المقترح والبالغة 75.8% وكما موضح بالجدول

Classification Results[a]

		y	Predicted Group Membership 0	1	Total
original	Count	0	48	12	60
		1	17	43	60
	%	0	80.0	20.0	100.0
		1	28.3	71.7	100.0

a. 75.8% of original grouped cases correctly classified.

وكانت دالة التصنيف لسنتين قبل الفشل وفقا للاتي:-

$$L = -1.4676 + 2.365X_2 - 8.149X_3 + 3.131X_5 + 0.356X_7$$

حيث :

X2= موجودات متداولة/إجمالي الموجودات ، X3= صافي الدخل / إجمالي الموجودات

X5= الأموال الخاصة / إجمالي الديون ، X7= أسهم / مبيعات

حيث ان L تمثل درجة الفشل.، فإذا كانت L>0 تصنف الشركة على إنها فاشلة ، وخلاف ذلك تشير إلى عدم الفشل. والجدول التالي يوضح المتغيرات الداخلة ومعنوية النموذج المقترح بالاعتماد على Ψ وعلى F وكالاتي:-

Variables Entered/Removed(a,b,c)

Step	Entered	Wilks' Lambda Statistic	df1	df2	df3	Exact F Statistic	df1	df2	Sig.
1	X5	0.715	1	1	118	46.938	1	118	.000
2	X3	0.562	2	1	118	45.672	2	117	.000
3	X2	0.528	3	1	118	34.620	3	116	.000
4	X7	0.494	4	1	118	29.395	4	115	.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

a Maximum number of steps is 18.

b Minimum partial F to enter is 3.84.

c Maximum partial F to remove is 2.71.

يلاحظ أهمية كل من المتغيرات الداخلة بالنموذج وفقا للجدول التالي :-

Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	x5	1.000	46.938	
2	x5	.997	38.125	.745
	x3	.997	32.055	.715
3	x5	.917	20.890	.623
	x3	.997	29.365	.661
	x2	.919	7.467	.562
4	x5	.463	28.411	.617
	x3	.979	31.781	.631
	x2	.887	9.841	.537
	x7	.448	7.711	.528

ويمكن ملاحظة نسب الخطأ في التصنيف للنموذج الثاني، إذ أن نسبة المفردات المصنفة بصورة صحيحة والتي تبين دقة التصنيف للنموذج المقترح بلغت 81.7% وكما موضح بالجدول :-

Classification Results[b]

		y	Predicted Group Membership		Total
			0	1	
Original	Count	0	48	12	60
		1	10	50	60
	%	0	80.0	20.0	100.0
		1	16.7	83.3	100.0

b. 81.7% of original grouped cases correctly classified.

وكانت دالة التصنيف لسنة واحدة قبل الفشل كالأتي

$$L = -0.5673 + 9.364X_3 + 0.551X_5$$

X3= صافي الدخل / إجمالي الموجودات ، X5= الأموال الخاصة / إجمالي الديون
L= درجة الفشل ، حيث إذا كانت L>0 تصنف الشركة على إنها فاشلة، وبالعكس فأنها ليست فاشلة. والجدول التالي يوضح المتغيرات الداخلة ومعنوية النموذج المقترح بالاعتماد على Ψ وعلى F وكالاتي:-

Variables Entered/Removed(a,b,c)

Step	Entered	Wilks' Lambda							
		Statistic	df1	df2	df3	Exact F			
						Statistic	Df1	df2	Sig.
1	x5	0.778	1	1	118	33.705	1	118	.000
2	x3	0.716	2	1	118	23.166	2	117	.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

a Maximum number of steps is 18.

b Minimum partial F to enter is 3.84.

c Maximum partial F to remove is 2.71.

يلاحظ أهمية كل من المتغيرات الداخلة بالنموذج وفقا للجدول التالي :-

Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	X5	1.000	33.705	
2	X5	.998	27.945	.887
	X3	.998	10.044	.778

وان نسبة الخطأ في التصنيف للنموذج الثالث الموضحة بالجدول أدناه إذ إن نسبة المفردات المصنفة بصورة صحيحة والتي تبين دقة التصنيف للنموذج المقترح بلغت 73.3%:-

Classification Results[a]

		V2	Predicted Group Membership 0	1	Total
Original	Count	0	49	11	60
		1	21	39	60
	%	0	81.7	18.3	100.0
		1	35.0	65.0	100.0

a. 73.3% of original grouped cases correctly classified.

2- الانحدار اللوجستي Logistic Regression (LR)

طبقت بيانات مع هذا المثال باستعمال برنامج SAS. المصنفات التي تم الحصول عليها من عينة 120 شركة، ولثلاث سنوات متتالية قبل الفشل وباستعمال LR. وحسب دالة الاحتمال:

$$P(F) = \frac{_e g(X)}{1 +_e g(X)}$$

لكل واحد من النماذج الثلاثة. حيث P(F) هي احتمالية الفشل Failure Probability وإذا كانت P(F)>0.5 تُصنف الشركة حينئذ على إنها فاشلة ، وعكس ذلك فإنها غير فاشلة ، ونلاحظ ان نموذج لثلاث سنوات قبل الفشل هو :

$$g(X) = -1.3287 + 1.4877X_4 + 1.6896X_5$$

حيث:-

X4= دخل قبل الفائدة والضرائب / اتفاقات مالية ، X5= الأموال الخاصة / إجمالي الديون

والجدول التالي يبين معنوية معلمات لمتغيرات الداخلة في النموذج كالأتي:

Variables in the Equation

	B	S.E.	Wald	Df	Sig.	Exp(B)
X5	1.689	.043	1.354	1	.024	5.417
X4	1.487	2.503	1.697	1	.019	4.426
Constant	-1.328	1.090	4.359	1	.037	9.744

وان دقة التصنيف باستخدام النموذج يمكن استنتاجها من - مصفوفة الخلط- والتي بلغت 60.8% وكما موضحة بالجدول الأتي:-

Classification Table(a)

Predicted Y	Observed Y 0	Observed Y 1	Percentage Correct
0	28	32	46.7
1	15	45	75.0
Overall Percentage			60.8

a The cut value is .500

أما نموذج الانحدار اللوجستي الخاص بسنتين قبل الفشل فكانت دالته كما موضح بالمعدلة التالية:

$$g(X) = -0.4999 - 13.589X_4 + 3.805X_5$$

X3= صافي الدخل / إجمالي الموجودات ، X5= الأموال الخاصة / إجمالي الديون

والجدول التالي يبين معنوية معلمات لمتغيرات الداخلة في النموذج كالأتي:

Variables in the Equation

		B	S.E.	Wald	Df	Sig.	Exp(B)
1	X5	3.805	.066	4.874	1	.027	44.961
	X3	-13.589	.536	17.922	1	.000	.123E-5
	Constant	-.499	3.396	17.515	1	.000	.606

كما أن دقة التصنيف باستخدام النموذج الثاني والتي بلغت 72.5% وان مصفوفة الخلـط كانـت كما موضحة بالجدول أدناه:-

Classification Table[a]

Observed		Predicted Y 0	Predicted Y 1	Percentage Correct
Y	0	38	22	63.3
	1	11	49	81.7
Overall Percentage				72.5

a. The cut value is .500

أما نموذج سنة واحدة قبل الفشل فقد كانت وفقا للصيغة التالية :-

$$g(x) = -1.2584 + 26.1304X_3 + 1.3535X_5$$

X3= صافي الدخل / إجمالي الموجودات ، X5= الأموال الخاصة / إجمالي الديون

والجدول التالي يبين معنوية معلمات لمتغيرات الداخلة في النموذج كالأتي

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
X5	1.354	.066	4.874	1	.027	3.871
X3	26.130	.536	17.922	1	.000	2.20E+011
Constant	-1.285	3.396	17.515	1	.000	.277

كما كانت دقة التصنيف باستخدام النموذج الثالث –سنة قبل الفـشل- 75% ومـصفوفة الخلـط موضحة كالأتي:-

Classification Table[a]

Observed		Predicted Y 0	Predicted Y 1	Percentage Correct
Y	0	46	14	76.7
	1	16	44	73.3
Overall Percentage				75.0

a. The cut value is .500

3- اشجار التصنيف (Classification Trees (CART

تم تصنيف المتغيرات إلى ثلاثة أصناف. وفيما يلي توضيحا لكيفية تنفيذ هذه العمليـة لكـل واحدة من هذه السينين الثلاثة. ولكل مجموعة من 60 شركة عاملة بصورة جيدة و60 شـركة فاشلة . وباستعمال التكرارات FREQUENCIES والوصـف DESCRIPTIVES مـن برنامج SPSS13 ، تم توزيعاتها وبالصورة التالية.

"*************** ثلاث سنوات قبل الفشل ***********************"

x1 (lowest through 0.30=1) (0.31 through 0.66=2) (0.67 through highest=3)
x2 (lowest through 0.63=1) (0.64 through 0.77=2) (0.78 through highest=3)
x3 (lowest through 0.02=1) (0.03 through 0.06=2) (0.07 through highest=3)
x4 (lowest through 0.21=1) (0.22 through 0.61=2) (0.62 through highest=3)
x5 (lowest through 0.24=1) (0.25 through 0.51=2) (0.52 through highest=3)
x6 (lowest through 0.32=1) (0.33 through 0.68=2) (0.69 through highest=3)
x7 (lowest through 0.11=1) (0.12 through 0.22=2) (0.23 through highest=3)
x8 (lowest through 0.21=1) (0.22 through 0.32=2) (0.33 through highest=3)
x9 (lowest through 0.07=1) (0.08 through 0.16=2) (0.17 through highest=3)

****************** "سنتين قبل الفشل"************************

x1 (lowest through 0.26=1) (0.27 through 0.62=2) (0.63 through highest=3)
x2 (lowest through 0.61=1) (0.62 through 0.74=2) (0.75 through highest=3)
x3 (lowest through 0.02=1) (0.03 through 0.06=2) (0.07 through highest=3)
x4 (lowest through 0.24=1) (0.25 through 0.49=2) (0.50 through highest=3)
x5 (lowest through 0.20=1) (0.21 through 0.43=2) (0.44 through highest=3)
x6 (lowest through 0.31=1) (0.32 through 0.59=2) (0.60 through highest=3)
x7 (lowest through 0.13=1) (0.14 through 0.24=2) (0.25 through highest=3)
x8 (lowest through 0.24=1) (0.25 through 0.36=2) (0.37 through highest=3)
x9 (lowest through 0.05=1) (0.06 through 0.12=2) (0.13 through highest=3)

******************* "سنه قبل الفشل"************************

x1 (lowest through 1.05=1) (1.06 through 1.46=2) (1.47 through highest=3)
x2 (lowest through 0.63=1) (0.64 through 0.77=2) (0.78 through highest=3)
x3 (lowest through -0.01=1) (0.0 through 0.04=2) (0.05 through highest=3)
x4 (lowest through 0.89=1) (0.90 through 1.89=2) (1.90 through highest=3)
x5 (lowest through 0.32=1) (0.33 through 0.85=2) (0.86 through highest=3)
x6 (lowest through 2.89=1) (2.90 through 5.89=2) (5.90 through highest=3)
x7 (lowest through 0.11=1) (0.12 through 0.22=2) (0.23 through highest=3)
x8 (lowest through 0.21=1) (0.22 through 0.32=2) (0.33 through highest=3)
x9 (lowest through -0.05=1) (-0.04 through 0.05=2) (0.06 through highest=3)

ولغرض تنفيذ طريقة الأشجار البيانية تم استخدام برنامج STATISTICA 6. وكانت المصنفات التي تم الحصول عليها من عينة من 120 شركة كمجموعة تدريب، لكل واحدة من ثلاث سنوات قبل الفشل، هي عبارة عن أشجار تصنيف ثنائيـة، اذ ان : F : تمثـل صـنف الفاشـلة (فاشلة) ، H : تمثل صنف غير الفاشلة (معافاة) ، FINAL NODE: هي عقـدة الورقـة إذ تنتهي العملية ، إذ يمثل الحرف F او H الفرد الذي يأتي في هذه العقدة والصنف المخصص هو فاشلة او غير فاشلة على التوالي. وتُحسب احتمالية الفشل بـ $\frac{H}{H+F}$ في حالة تصنيف الشركة على انهاغير فاشلة، وعند الوصول الى عقدة الورقة ، بـ $\frac{F}{H+F}$ في حالة تـصنيف الـشركة على انها فاشلة .

ويمكن تلخيص النتائج التحليلات السابقة وفقا للجدول التالي :

النموذج / عدد السنوات قبل الفشل	DA	LR	CART
1	73.3	75	80
2	81.7	72.5	73.3
3	75.8	60.8	59.3

ويلاحظ تجانس النسب المئوية للتصنيف الصحيح مع الزمن قبل الفشل وقد يشذ عن هذه الحالة التصنيف التمييزي إذ يبدي قيمة متميزة عند السنتين إذ بلغت 81.7 % وقد أبدى تغايرا مع الزمن وقد يكون السبب في ذلك تجانس البيانات ضمن فتراتها، أماّ أفضل أداء فكان عند استخدام أسلوب التصنيف الشجيري، وكانت الطرق المعلميه متقاربة في أدائها وإنْ اظهر أسلوب الانحدار اللوجستي استقرارا أفضل ودقه أعلى في النموذج الذي يمثل سنة واحدة قبل الفشل إذ بلغ 75 % بالمقارنة مع DA إذ بلغ 73% تقريبا . في حين اظهر طريقة DA هي الافضل بالنسبة للنماذج التي تمثل سنتين وثلاث سنوات قبل الفشل. وبصورة عامه يلاحظ ارتفاع جودة التصنيف لنتائج السنة الواحدة قبل الفشل ولعل ذلك يعود إلى عدم تعرض بيانات الظاهرة المدروسة لسنة واحدة إلى تأثيرات عرضيه تسهم في ذبذبة البيانات مما يفقدها اتجاهها العام الأمر الشائع في البيانات المالية والظواهر الاقتصادية بصورة عامة .

8- الاستنتاجات

1- نرى ان التحليل المميز ونماذج الانحداراللوجيستي لهما المتغيرات نفسها (النسب)- بصورة عامة حيث اشتملت بالسنة الاولى قبل الفشل على المتغير x_5,x_3 والذي يعكس اهمية تلك المتغيرات، اما لسنتين قبل الفشل فقد كانت النتائج للنموذجين المعلميين متقاربة ايضّا اذ اشتمل النموذجين على المتغييرين x_3,x_5 وكانت نسبة تصنيف الـ LR ادنى من نسبة تصنيف نموذج DA الذي احتوى على اربعة متغيرات بلغت دقة تصنيفه 81% في حين احتوى الاخير على متغيرين فقط وللسنين الثلاثه قبل الفشل واحتوى النموذجان على ذات المتغيرات x_5,x_4 وان النسب المؤيه للمفردات المصنفه تصنيفا جيدا تكون تقريبا متشابهة، وفي حالة استخدام النموذج اللوجستي تكون المصنفات افضل قليلاً.

2- من السهل ملاحظة اشجار التصنيف، فكلما كان تاريخ الفشل قريبا كان عمق الشجرة قليلا، والذي يمكن ان يفهم على انه كلما كان تاريخ الفشل قريبا قلت ضرورة فحص المتغيرات، كلما تسارع الوصول الى خاتمة حول مصير المفردات في مجموعات الفحص ولعل الفشل عندما يكون قريبا تقل الحاجة الى القوانين المراد تفحصها لتحديد "مصير" الشركة وان ادءها كان افضل في حالة الثلاث سنوات مقارنة مع اداء بقية النماذج للفترة نفسها وربما يكون ذلك مبنياً على الاستراتيجية التي تقوم عليها هذه الطريقة إذ لا توجد عمليات حسابية مما يساعد على التركيز على الخصائص الفعلية للمفردات إذ تقوم العملية على اساس استخراج القواعد العلائقيه بين الخواص (المتغيرات) وحسب فئاتها.

9- التوصيات

1- استخدام اسلوب الاشجار البيانية في التنبؤ بفشل الشركات بصورة عامه.
2- استخدام الاساليب المعلمية في الفترات المتوسطة سنتين كما يمكن استخدام الاساليب الاخرى (التكراريه والشبكات العصبية لنفس الغرض).
3- المتغيرات التي يمكن الاهتمام بها عند تقويم اداء الشركات هو x_5,x_3 بصورة عامه وكما موضح بالاستنتاجات .
4- تطبيق النماذج على الشركات العراقية ولسنوات اطول وقطاعات اخرى.

المصادر العربية:-

1- موراي . ر.شبيغل 1978-سلسلة ملخصات شوم – نظريات ومسائل في الإحصاء.

الرسائل والاطاريح الجامعية:-

2- سعيد، ميعاد مسعود (1986)، استخدام الدالة المميزة لدراسة العوامل المؤثرة على المستوى العلمي في معاهد المعلمين والمعلمات في محافظة بغداد كلية الإدارة والاقتصاد، جامعة بغداد.
3- يعقوب، هيثم منير (1995) استخدام التحليل المميز المتعدد في تخصيص العوامل المؤثرة في التصنيف السريري لمرضى القلب، رسالة مقدمة إلى مجلس كلية الإدارة والاقتصاد في الجامعة المستنصريه.

المصادر الاجنبية:-

4- Anderson J.A (1982) Logistic Discriminate, hand book of statistical Anderson J.Aol,2,north –Holland and publishing company.

5- Anderson J.A, and V. Elia. Penalized Maximum Likelihood estimation in Logistic regression and discriminate.

6- Argenti, J. (1976). *Corporate Collapse: the Causes and Symptoms*. McGraw- Hill. London.

7- Blum, M. (1974). Failing Company Discriminate Analysis. *Journal of Accounting Research*, 1-23.

8- Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). *Classification and Regression Trees.* Monterey, CA: Wadswooeth and Brooks.

9- Deakin, E.B. (1972). A Discriminate Analysis of Predictors of Business Failure. *Journal of Accounting Research*, 167-179.

10- Hosmer, D. W. and Lemeshow, S. (1989). *Applied Logistic Regression*. Wiley Series in Probability and Mathematical Statistics.

11- Lauritzen, S.L. (1996). *Graphical models*. Oxford Science Publications.

12- SAS Institute Inc. (1993). *SAS Language: Reference, Version 6*, SAS Institute Inc.