



الجمهورية العربية السورية
وزارة التعليم العالي
جامعة حلب
كلية الاقتصاد
قسم الاحصاء ونظم المعلومات

بناء شجرة القرارات للتنبؤ بحجم المبيعات باستخدام تقنيات التنقيب في البيانات

(دراسة تطبيقية على هايبر ماركت)

**Building a Decision Tree to Predict the Volume of Sales by
Using Data Mining Techniques**
(An Applied Study on a Hypermarket)

دراسة أعدت لنيل درجة الماجستير في نظم المعلومات الإدارية

إعداد الطالب
أحمد محمد وهبي عقاد

كلية الاقتصاد - جامعة حلب

1439 هـ - 2017 م



الجمهورية العربية السورية
وزارة التعليم العالي
جامعة حلب
كلية الاقتصاد
قسم الاحصاء ونظم المعلومات

بناء شجرة القرارات للتنبؤ بحجم المبيعات باستخدام تقنيات التنقيب في البيانات

(دراسة تطبيقية على هايبر ماركت)

**Building a Decision Tree to Predict the Volume of Sales by
Using Data Mining Techniques**

(An Applied Study on a Hypermarket)

دراسة أعدت لنيل درجة الماجستير في نظم المعلومات الإدارية

إعداد الطالب

أحمد محمد وهبي عقاد

إشراف الدكتور

غسان ساكت

أستاذ مساعد في قسم الاحصاء ونظم المعلومات

كلية الاقتصاد - جامعة حلب

1439 هـ - 2017 م

مَهْمَةُ سَيِّدِ الْجَمَاعَةِ

أصرح بأن هذا البحث عنوان :

بناء شجرة القرارات للتنبؤ بحجم المبيعات باستخدام تقنيات التقيب في البيانات (دراسة تطبيقية على هايبر ماركت)

لم يسبق أن قيل للحصول على أية شهادة، ولا هو مقدم حالياً للحصول على شهادة أخرى.

المرشح

أحمد محمد وهبي عقاد

نوقشت هذه الرسالة وأجيزت بتاريخ 23 تشرين الثاني 2017

لجنة المناقشة والحكم

عضوًأ

عضوًأ ومشرفاً

رئيسًأ

أ.د. محمد ضاهر

أ.د. غسان ساكت

أ.د. ابراهيم النائب

لجنة المناقشة والحكم

لجنة الحكم على رسالة الماجستير في نظم المعلومات الإدارية بقسم الاحصاء ونظم

المعلومات المقدمة من الطالب :

أحمد محمد وهبى عقاد

عنوان

بناء شجرة القرارات للتنبؤ بحجم المبيعات باستخدام تقنيات التقييم في البيانات

(دراسة تطبيقية على هايبر ماركت)

الدكتور: إبراهيم نائب الأستاذ في قسم الاحصاء ونظم المعلومات رئيساً

كلية الاقتصاد-جامعة حلب

الدكتور: غسان ساكت الأستاذ المساعد في قسم الاحصاء ونظم المعلومات عضواً ومشرفاً

كلية الاقتصاد-جامعة حلب

الدكتور: محمد ضاهر الأستاذ المساعد في قسم الاحصاء ونظم المعلومات عضواً

كلية الاقتصاد-جامعة حلب

سال٢٠١٩
جامعة حلب

نشهد بأن العمل المقدم في الرسالة هو نتيبة بحث علمي قام به المرشح أحمد محمد وهبي عقاد بإشراف الدكتور غسان ساكت أستاذ مساعد في قسم الإحصاء ونظم المعلومات من كلية الاقتصاد جامعة حلب، وأية مراجع أخرى ذكرت في هذا العمل موثقة في نص الرسالة وحسب ورودها في النص.

المرشح

أحمد محمد وهبي عقاد

المشرف

أ.د غسان ساكت

اللهم قدراك
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

إِنَّمَا يُنَزَّلُ مِنْ رَبِّكَ
الرُّوحُ مِنْ أَنْفُسِ الْأَنْوَارِ

”اللهم جل جلاله“

إِنَّمَا يُعْلَمُ الْأَوَّلُ... وَالصَّاعِدُونَ إِلَيْنَا... سَيِّدُنَا

”رسول الله محمد صلى الله عليه وسلم“

إِنَّمَا يُعْلَمُ كُلُّ مَا يَعْلَمُ حَتَّىٰ تَكُونَ الرُّرُوعُ إِنَّمَا يَعْجَبُ بِرَرَاعَ نَبَاتَ بِفَضْلِ اللَّهِ

”أَبْيَ أَمْرَ اللَّهِ فِي حُمْرَهٖ وَعَافِيَّهٖ“

إِنَّمَا يُعْلَمُ الْخَيَّا وَالْعَطَاءُ وَالْحُبُّ ...

إِنَّمَا يُعْلَمُ مَنَّاكَ وَزَمَانَكَ ...

”أَبْيَ أَمْرَ اللَّهِ فِي حُمْرَهٖ وَعَافِيَّهٖ“

إِنَّمَا يُعْلَمُ أَيَّامُ الْغَفْوَةِ وَالْتَّبَابِ وَسَنَدِيَ فِي هَذِهِ الْجَمَاهِيرِ

”أَخْوَانِي أَمَانِي وَأَسْمَاءَ“

إِنَّمَا سَانَرْنِي وَلَهُمْنِي فِي حَمْلِي وَعَوْرِي

”الغالبة ما وونا عجبوا ز“

إِنَّمَا تَحْلُّ بِالْإِخْرَاءِ وَعَيْرُوا بِالْمَوْفَاءِ وَالْعَطَاءِ إِنَّمَا يَنْبَغِي الصَّرْقُ لِلصَّارِفِ

”أَصْدِرْفَانِي“

إِنَّمَا يَلْدِرِي الْجَيْسِ سُورَةٌ ... أَهْدِرِي غَرَّةً جَهْرِي الْمَنْوَاضِعَ

شكر وامتنان

الحمد والشكر لله الذي منَّ علي بإتمام هذا البحث المتواضع، وأرجو من الله أن يتقبله مني و يجعله خالصاً لوجه الكريم، وإيماناً بفضل الاعتراف بالجميل، فإني أتقدم بالشكر لوالدي لما قدماه لي من دعاء وتشجيع، ولكل من ساعد في إنجاز الرسالة وأخص بالذكر:

أستاذي ومشرفي الفاضل الدكتور غسان ساكت الأستاذ المساعد في قسم الاحصاء ونظم المعلومات في كلية الاقتصاد بجامعة حلب، على قبوله الاشراف وعلى ما منحني إياه من نصح وإرشاد، جزاه الله خيراً وأنعم عليه بنعمة الصحة والسعادة، وجعل عمله هذا في ميزان حسناته.

كما أتقدم بالشكر الجليل وعظيم الاحترام للدكتور الفاضل إبراهيم نائب الأستاذ في قسم الاحصاء ونظم المعلومات في كلية الاقتصاد بجامعة حلب على تفضله بقبول مناقشة هذا البحث، جزاه الله خيراً وأنعم عليه بنعمة الصحة والسعادة، وجعل عمله هذا في ميزان حسناته.

كما أتقدم بالشكر الجليل وعظيم الاحترام للدكتور الفاضل محمد ضاهر الأستاذ المساعد في قسم الاحصاء ونظم المعلومات في كلية الاقتصاد بجامعة حلب على تفضله بقبول مناقشة هذا البحث، جزاه الله خيراً وأنعم عليه بنعمة الصحة والسعادة ، وجعل عمله هذا في ميزان حسناته.

كما أتوجه بخالص الشكر والامتنان إلى جامعة حلب رئاسةً، وعمادةً وأساتذةً وإداريين، والشكر الجليل لكلية الاقتصاد وأعضاء الهيئة التدريسية في قسم الاحصاء ونظم المعلومات وأخص بالذكر رئيس القسم الدكتور عبد الرحمن عبيد، والدكتور ياسر الموسى

والدكتورة أميرة عبيدو، والدكتورة نور حميدي على توجيهاتهم القيمة جزاهم الله خيراً وأنعم عليهم بنعمة الصحة والسعادة الدائمة، وجعل عملهم هذا في ميزان حسناتهم.

وختاماً اللهم إن كنت قد أصبت في مقصدي فبنعمة منك، وإن كنت قد أخافت فهو من عندي، وأسالك اللهم أن يكون عملي هذا وعمل كل من أعاذني عليه لنفع بلدي سوريا، والحمد والشكر لله رب العالمين.



هدف البحث الى دراسة علاقات المواد مع بعضها البعض في الهايبر ماركت من خلال تحليل سلة التسوق باستخدام خوارزمية تحليل قواعد الاقتران عبر برنامج SQL Sever Business Intelligence، وتم الحصول من هذا التحليل على قواعد وعلاقات بين المواد، استطعنا من خلالها بناء نموذج للتنبؤ بحجم المبيعات بعد اضافة متغيرات مستقلة إضافية وتحليل هذه البيانات التاريخية على مر ثلاث سنوات باستخدام خوارزمية M5P بالاستعانة ببرنامج Weka، وقد تم استخدام خاصية التقسيم الفئوي Percentage Split في عملية تقسيم البيانات الى بيانات تدريب وبيانات اختبار وتم الحصول على شجرة القرارات التي تدعم عملية التنبؤ.

وتم أيضاً مقارنة التنبؤ بحجم المبيعات بين أشجار القرار وبين الطرق الإحصائية التقليدية (معادلات الانحدار) وتم التوصل الى النتائج التي تثبت امكانية الاستفادة من النتائج المتولدة من تحليل قواعد الاقتران في عملية بناء شجرة القرارات.

الكلمات المفتاحية: التنبؤ بحجم المبيعات، تحليل سلة التسوق، قواعد الاقتران، أشجار القرار، M5P.



 كلية الاقتصاد والعلوم السياسية
 جامعة القاهرة

رقم الصفحة	العنوان
الإطار العام للبحث	
أ	مقدمة
ب	مشكلة البحث
ب	أهداف البحث
ب	أهمية البحث
ت	منهجية البحث
ت	حدود البحث
ت	الدراسات السابقة
الفصل الأول: التنبؤ بالمبيعات	
2	المبحث الأول التنبؤ بالمبيعات وطرقه
2	1.1.1 مقدمة
3	2.1.1 مفهوم التنبؤ
4	3.1.1 تعريف التنبؤ بالمبيعات
5	4.1.1 أهداف التنبؤ بالمبيعات
5	5.1.1 أهمية التنبؤ بالمبيعات وعلاقته بتحطيط المبيعات
7	6.1.1 أبعاد عملية التنبؤ بالمبيعات
8	7.1.1 العوامل المؤثرة في دقة التنبؤ بالمبيعات
9	1.7.1.1 العوامل الخارجية
9	2.7.1.1 العوامل الداخلية
10	8.1.1 مدة التنبؤ
11	9.1.1 خطوات التنبؤ

12	10.1.1 طرق التنبؤ بالمبيعات
12	1.10.1.1 الطرق النوعية (الوصفية)
12	1.1.10.1.1 الرأي الجماعي
12	2.1.10.1.1 أسلوب لجنة الخبراء
13	3.1.10.1.1 طريقة دلفي
13	2.10.1.1 النماذج الكمية
13	1.2.10.1.1 طريقة المتوسط البسيط
14	2.2.10.1.1 طريقة المتوسطات المتحركة البسيطة
14	3.2.10.1.1 طريقة المتوسطات المتحركة المرجحة
14	4.2.10.1.1 طريقة التمهيد الأسوي البسيط
14	5.2.10.1.1 طريقة تحليل السلالس الزمنية
15	3.10.1.1 الطرق التقنية
15	1.3.10.1.1 الشبكات العصبية
15	2.3.10.1.1 أشجار القرار
17	المبحث الثاني: الموازنة التقديرية للمبيعات
17	1.2.1 مقدمة
17	2.2.1 مفهوم الموازنة التقديرية
18	3.2.1 أهمية الموازنة التقديرية للمبيعات
19	4.2.1 مراحل اعداد الموازنة التقديرية للمبيعات
20	5.2.1 مدة الموازنة التقديرية للمبيعات
الفصل الثاني: التقييم في البيانات	
23	المبحث الأول: عناصر التقييم في البيانات وتطبيقاتها
23	1.1.2 مقدمة
23	2.1.2 مفهوم التقييم في البيانات
24	3.1.2 صعوبات التقييم في البيانات
25	4.1.2 مهام التقييم في البيانات

26	5.1.2 منهجيات التقريب في البيانات
27	6.1.2 تطبيقات التقريب في البيانات
28	7.1.2 خطوات التقريب في البيانات
30	المبحث الثاني: تحليل سلة التسوق وقواعد الارتباط
30	1.2.2 مقدمة
30	2.2.2 تعريف قواعد الاقتران
30	3.2.2 أنواع قواعد الاقتران
31	4.2.2 تطبيقات قواعد الاقتران
31	5.2.2 مفاهيم قواعد الاقتران
32	6.2.2 خوارزميات قواعد الاقتران
32	1.6.2.2 Apriori خوارزمية
33	2.6.2.2 FP Tree خوارزمية
35	المبحث الثالث أشجار القرارات
35	1.3.2 مقدمة
35	2.3.2 مفهوم أشجار القرارات
37	3.3.2 مميزات وعيوب أشجار القرارات
38	4.3.2 تطبيقات أشجار القرارات
38	5.3.2 استقراء شجرة القرارات
39	6.3.2 خوارزميات أشجار القرارات
41	1.6.3.2 M5 خوارزمية
43	7.3.2 نقاط القوة والضعف لأشجار القرارات
الفصل الثالث دراسة حالة التطبيقية	
47	المبحث الأول دراسة سلة التسوق
47	1.1.3 مفهوم الهايبر ماركت
47	2.1.3 توثيق البيانات للتحليل
52	3.1.3 فوائد نتائج قواعد الاقتران

53	4.1.3 نتائج البحث الأول
54	المبحث الثاني: التنبؤ بحجم المبيعات بالاعتماد على بيانات سلة التسوق
54	1.2.3 مقدمة
54	2.2.3 المتغيرات الداخلة في التنبؤ
54	3.2.3 تهيئة البيانات
55	4.2.3 تخفيض عدد المتغيرات باستخدام تحليل قواعد الاقتران
60	5.2.3 تحليل البيانات عن طريق برنامج Weka
61	6.2.3 بناء نموذج شجرة قرارات وتقديرها
66	7.2.3 تقدير النموذج
67	8.2.3 هيكلية شجرة القرارات
69	9.2.3 استخدام طريقة الانحدار المتعدد في التنبؤ
72	الفصل الرابع النتائج والتوصيات
72	1. النتائج
73	2. التوصيات
74	المراجع
80	الملحق

رقم الصفحة	الجدول
47	الجدول 1: جزء من بيانات التحليل قبل التهيئة
49	الجدول 2: جزء من بيانات التحليل بعد التهيئة
51	الجدول 3: أبرز قواعد الاقتران
56	الجدول 4: جزء من البيانات المهمة للتتبؤ بأحجام المبيعات الأسبوعية
57	الجدول 5: جزء من النتائج الظاهرة في تحليل معامل بيرسون
59	الجدول 6: جزء من البيانات بعد التهيئة للتتبؤ بأحجام المبيعات الأسبوعية
60	الجدول 7: جزء من البيانات الأسبوعية للمبيعات التي سيتم تحليلها
69	الجدول 8: ملخص تحليل النماذج
70	الجدول 9: جدول تحليل التباين
70	الجدول 10: تحليل معنوية المتغيرات

فَمَا يَرَى إِلَّا مَا أَنْشَأَ اللَّهُ
يَسِيرٌ سَرِيرًا

العنوان	رقم الصفحة
الشكل 1: العلاقة بين التبؤ بالمبيعات وتخطيط المبيعات	6
الشكل 2: مهام التقييب في البيانات	25
الشكل 3: خطوات التقييب في البيانات	29
الشكل 4: مكونات شجرة القرار	36
الشكل 5: ارتباط عقد الشجرة	36
الشكل 6: آلية بناء شجرة القرارات بخوارزمية M5P	42
الشكل 7: مخطط الدراسة التطبيقية	46
الشكل 8: الدعم عند مستويات عتبة متعددة	50
الشكل 9: الثقة عند مستويات عتبة متعددة	50
الشكل 10: جزء من شجرة القرارات قبل التشذيب	62
الشكل 11: شجرة النماذج بعد التشذيب	63
الشكل 12: جزء من شجرة النماذج بعد التشذيب	68

الإطار العام للبحث

General Framework of Research

١. المقدمة:

حظيت دراسة السلوك الإنساني بأهمية كبيرة منذ قديم الزمان وذلك لأهميته ومحاولة التعرف عليه والوقوف على العوامل التي تؤثر فيه، وذلك لأن سلوك الأفراد يتتأثر بعوامل عديدة يتقاوم تأثيرها باختلاف هذه العوامل، حيث هناك اختلاف واضح في سلوك الأفراد وذلك نتيجة لاختلاف الدول والمناطق والأجناس والعادات والتقاليد والقوانين والأعراف والحضارات وغيرها من العوامل. وبما أن سلوك المستهلك يمثل أحد جوانب هذا السلوك فإن دراسته ومحاولة التعرف عليه حظيت باهتمام كبير جداً من مختلف المدارس الاقتصادية والسلوكية وأصبح في الوقت الحاضر نقطة مهمة وأساسية وتحتل المكانة الأولى بالنسبة للمنظمات وذلك نتيجة للتقدم العلمي والتكنولوجي والذي أدى إلى زيادة أعداد المنتجات المطروحة في السوق وتتنوعها، وجود منتجات بديلة لها ممكن أن تحل محلها وتعطي خدمات ومنافع مشابهة أو مقاربة، واستداد المنافسة وتطور وتتنوع أدوات المستهلكين بالشكل الذي جعل سلوك المستهلك نقطة البداية والنهاية لمختلف الأنشطة الإنتاجية والتسويقية والأنشطة الأخرى في المنظمات، مما دفع المختصين لدراسة سلوك المستهلك ومحاولة التعرف على حاجاته ورغباته ومحاولة تلبيتها بهدف تحقيق الرضا والإشباع لديه ودفعه إلى اتخاذ قرار الشراء. ويرتبط نجاح هذه المنظمات في الوقت الحاضر بمدى قدرتها على تلبية هذه الرغبات والاحتاجات وتحقيق الرضا للمستهلك.

إن دراسة هذا السلوك يمكن الإدارة المالية في المنظمة من تقدير حجم المبيعات المتوقع في حال قيامنا بتقديم المنتجات وفق رغبات الزبون ويؤدي بالضرورة إلى التحكم في زيادة حجم المبيعات والوصول بالمنظمة إلى هدفها الأساسي وهو زيادة الربحية وتحقيق المنافسة مع قرائها في السوق.

سنتناول في هذه الدراسة التنبؤ بحجم المبيعات في أحد الهايبر ماركات السورية بناءً على القدرة على دراسة السلوك الشرائي للمستهلك والاستفادة من هذا السلوك من أجل عملية زيادة المبيعات الإجمالية للمنشأة وذلك وفق تقنيات حديثة في التقريب والتنبؤ.

2. مشكلة البحث:

تتلخص مشكلة البحث في وجود كم هائل من البيانات والتي تؤدي إلى:

- عدم قدرة العنصر البشري الخبير، أو طرق التحليل التقليدي في بعض الأحيان، على ملاحظة مدى ارتباط المواد مع بعضها البعض.
- قلة فعالية الطرق المستخدمة في معرفة أنواع المستهلكين وتوجهاتهم وتوظيف هذه المعرفة في زيادة المبيعات.
- عدم كفاءة الطرق المستخدمة للاستفادة من الكم الهائل من البيانات الناتجة عن نقاط البيع في دراسات التنبؤ بالمبيعات.

3. أهداف البحث:

- الاستفادة من الكم الهائل من بيانات فواتير نقاط البيع من أجل معرفة علاقات المواد مع بعضها البعض.
- تحليل البيانات باستخدام خوارزميات قواعد الاقتران للحصول على معلومات مفيدة والتي من الصعب الحصول عليها باستخدام الطرق التقليدية.
- العمل على التأثير في آراء الزبائن من خلال تقديم أو ترتيب منتجات جديدة وفق دراسة اتجاهات الزبائن والمنتجات وفوائير نقاط البيع التي يقوم بها الزبائن في قاعدة البيانات.
- بناء نموذج شجرة قرارات يدعم عملية التنبؤ بالمبيعات بالاعتماد على نتائج تحليل سلة التسوق.

4. أهمية البحث:

تأتي أهمية البحث من خلال النقاط التالية:

- تسمح دراسات تحليل سلة التسوق بتزويد متخذ القرار بكافة المعلومات التي تساعده على الاختيار الأمثل لسلع معينة وعلامات تجارية محددة من بين مختلف البائعين المعروضة بهدف المحافظة على رصيد المخزون من هذه البضاعة لتلبية احتياجات الزبائن.
- استخدام تقنيات حديثة تتمثل بالتنقيب في البيانات لاستخلاص المعرفة من بيانات الهابير ماركت والتي من الصعب استخلاصها باستخدام الطرق التقليدية.
- يساعد التنبؤ بالمبيعات في إعداد برنامج تسويقي فعلي متكامل للمنشأة أو على مستوى منتج أو مجموعة منتجات.

5. منهجية البحث:

حتى يتم التوصل بالبحث إلى الهدف المنشود سيتم الاعتماد على:

- **الجانب النظري:** المنهج الوصفي التحليلي وذلك بدراسة واستعراض النواحي المتعلقة بالتبؤ بحجم المبيعات وطرقه من خلال الأبحاث والمراجع العلمية وطرق استكشافه باستخدام تقنيات التقريب في البيانات حيث سيتم عرض قواعد الاقتران وأشجار القرارات والخوارزميات الخاصة بكل منها.
- **الجانب العملي:** حيث سيتم تحليل البيانات الخاصة بسلة التسوق باستخدام اداة SQL Server عن طريق خوارزميات قواعد الاقتران (Association Rules) واستخدام Business Intelligence النتائج التي سيتم الحصول عليها في عملية التنبؤ بالمبيعات باستخدام برنامج Weka، بعد التأكد من فائدة البيانات وتأثيرها باستخدام البرنامج الاحصائي SPSS.

6. حدود البحث:

تنقسم حدود البحث إلى:

- **الحدود المكانية:** وتشمل بيانات أحد الهايبر ماركت في مدينة حلب.
- **الحدود الزمنية:** وتشمل البيانات التاريخية لحجم مبيعات المواد في الهايبر ماركت من الفترة الممتدة من 2008 / ولغاية 2010 .

7. الدراسات السابقة:

- دراسة (الجريسي خالد، 2010) بعنوان:

"**سلوك المستهلك دراسة تحليلية للقرارات الشرائية للأسرة السعودية**"

مؤسسة الجريسي للتوزيع والإعلان، الرياض، السعودية.

يتعلق هذا البحث بالجوانب المتعددة لسلوك المستهلك، بالإضافة إلى دراسة قرار الشراء الاستهلاكي وأهم العوامل المرتبطة والمؤثرة فيه، والمدخل إلى ذلك هو مدخل منهجي يستند إلى التطبيق العلمي، وقد اختلف هذا البحث عن غيره من البحوث في المجال المنهجي المتكامل لكل من المفاهيم والأبعاد المتعددة للتسويق، إضافة إلى السلوك الاستهلاكي في آن معاً، وذلك من خلال ربط منطقي يستند إلى الممارسة العملية في هذا المجال.

وقد قام بدراسة قرار شراء جهاز حاسوب في الأسرة السعودية واستخدم في ذلك الطرق الإحصائية التقليدية في التحليل وقد توصل إلى أن هناك ميل إلى التجديد في الأسرة السعودية، أيضاً ثبت بأن ميل أي أسرة للتجدد وسرعة تبني المنتجات الجديدة يتأثر بالخلفية демографية لتلك الأسرة ومعدل دخل هذه الأسرة.

- دراسة (فاتح لقرقي، 2014) بعنوان:

"جودة نماذج السلسل الزمنية الموسمية المختلطة SARIMA في التنبؤ بالمبيعات"

دراسة حالة مؤسسة مطاحن جديع بتقررت

تعتبر نماذج السلسل الزمنية الموسمية المختلطة SARIMA من أهم نماذج السلسل الزمنية العشوائية الخطية فهي تأخذ بعين الاعتبار عدم الاستقرارية والتغيرات الفصلية في السلسلة الزمنية، حيث قام الباحث باختبار دقة هذه النماذج في تقدير سلسلة المبيعات الشهرية لمؤسسة مطاحن جديع من منتج الفرينة والنخالة في الفترة الممتدة من كانون الثاني 2008 إلى كانون الأول 2012، ومن ثم اختبار دقتها في التنبؤ بالمبيعات المستقبلية للتسعة أشهر الأولى من سنة 2013، ليتم فيما بعد مقارنتها بالقيم الفعلية واستنتاج مدى دقة نماذج SARIMA في مجال المبيعات. حيث أظهرت النتائج تقارب بين القيم الفعلية للمبيعات والقيم التي تم التنبؤ بها، وأنها تقع كلها دون استثناء داخل مجال التنبؤ، ليتم فيما بعد استخلاص دقة هذه النماذج وأنها يمكن أن يعود عليها في مجال التنبؤ بالمبيعات المستقبلية ل المؤسسة.

الدراسات الأجنبية:

- دراسة (Madani S., 2009) بعنوان:

"mining changes in customer purchasing behavior"

"التنقيب عن التغيير في سلوك شراء العملاء"

Master thesis continuation course marketing and e-commerce, Lulea University of technology.

قام **Madani** بدراسة سلوك المستهلك من خلال تحليل بيانات الزبائن وسلوك شرائهم عن طريق خوارزميات قواعد الارتباط حيث تم استخدام الخوارزمية Apriori Algorithm لتحليل بيانات شركة Kalleh للصناعة والتوزيع. وقد تم اعتماد مقياسين لقياس التغيير في السلوك وهما التشابه وعدم التوقع Similarity and Unexpectedness حيث تم تحليل درجة التشابه بين أنماط في فترات زمنية مختلفة، وقد تم اعتماد 10% لكل من عتبات الدعم والثقة، وقد أظهرت النتائج أن 6.5% من السلوك قد تم تطويره والحصول على مزيد من المعلومات حول سلوك المستهلك وهذا ساعد مدير التسويق بالتخطيط للاستجابة بدقة وفي الوقت المناسب للتغيرات في السوق.

- دراسة (Svetina M., & Zupancic J., 2015) بعنوان:

"How to Increase Sales in Retail with Market Basket Analysis"

"كيف نزيد المبيعات في محلات البيع بالتجزئة من خلال تحليل سلة التسوق"

University of Maribor, Kranj, Slovenia.

قام **Zupancic و Svetina** بدراسة مدى تأثير دعم عملية التسويق على زيادة المبيعات وذلك من خلال تحليل بيانات سلة التسوق. تمت الدراسة على بيانات شركة Merkur وهي شركة لتجارة القطع المنزلي، حيث قامت هذه الدراسة بتحديد قواعد الاقتران ومعدل العائد من بيع المواد الداخلة في هذه القواعد وتم الحصول على مواد ذات صلة مع مواد أخرى من أجل وضعها في حملات ترويجية للشركة وإبراز مدى مساهمة هذه المواد في دعم عمليات التسويق وزيادة المبيعات.

- دراسة (Gancheva V., 2013) بعنوان:

"Market basket analysis of beauty products"

"تحليل سلة التسوق لمنتجات التجميل "

Master of Science in Economics and Business, Erasmus University Rotterdam.

قام **Gancheva** بدراسة مدى قدرة تحليل بيانات سلة التسوق لشركة مواد تجميل في بلغاريا على دعم قرارات التسويق وإدارة المخزون، يحتوي المتجر على مواد تجميل محلية الصنع وثلاث أصناف عالمية (Brands) أخرى، حيث تم في عملية التحليل تجميع المواد تحت أصناف رئيسية محددة من أجل تقليل عدد المواد، وقد استخدم في التحليل أداة SQL Server Business Intelligence، وقد توصل الباحث إلى علاقات الأصناف بعضها ببعض وقام بتقديم توصيات في أي فصول من السنة يجب على المتجر زيادة مخزونهم من أصناف محددة.

- دراسة (Khalilzadeh N., 2008) بعنوان:

"Sales Prediction for Pharmaceutical Distribution Companies -A Data Mining Based Approach"

"التنبؤ بالمبيعات لشركات توزيع الأدوية اعتماداً على طريقة التنقيب في البيانات "

Master thesis continuation course marketing and e-commerce, Lulea University of technology.

كثير من شركات توزيع الأدوية تملك أهداف ربحية مثل الشركات العادي، إضافة للخدمات الإنسانية التي تقدمها، أي أنها تسعى إلى زيادة أرباحها وتقليل تكاليفها حيث تعاني هذه الشركات من مشكلة عجز أو فائض في المخزون من العقاقير وبالتالي خسارة جزء من الزبائن في حال العجز في المخزون حيث يحتاج حل هذه المشكلة إلى عملية التنبؤ بالمبيعات من أجل تلبية المتطلبات من العقاقير.

احتاج Khalilzadeh إلى بيانات سابقة عن مبيعات إحدى شركات الأدوية الإيرانية (Pakhsh Hejrat) من العقارات حيث تم جمع مبيعات ثلاثة سنوات سابقة وتم استخدام الشبكات العصبية كأداة رئيسية في عملية التنبؤ وقد تم بناء نموذج للتنبؤ بثلاث طرق مختلفة:

١. طريقة ARIMA للتنبؤ بالسلسل الزمنية.

٢. طريقة الشبكات العصبية الهجينة للتنبؤ بالسلسل الزمنية بواسطة البيانات الماضية لكل عقار.

٣. طريقة الشبكات العصبية الهجينة للتنبؤ بالسلسل الزمنية بواسطة البيانات الماضية لكل عقار وبواسطة جزء من البيانات الماضية للعقاقير المرتبطة معه.

وقد أظهرت النتائج أن النموذج الثالث هو أدق نموذج للتنبؤ من بين النماذجين الأول والثاني وعملية الاعتماد على العقاقير المرتبطة له أثر كبير على دقة النموذج.

- دراسة (Rajcan S., 2013) بعنوان:

"Magazine sales prediction"

"التنبؤ بمبيعات المجلة"

Master thesis, Charles University in Prague

يواجه العديد من ناشري المجلات مشكلة التنبؤ بمنتجاتهم، حيث أن كثير من هذه التنبؤات هي عبارة عن تخمينات من موظفين غالباً ما تكون خاطئة، والقرار الخاطئ بكل الحالتين إن كان بزيادة عدد المطبوعات أو تخفيضها وله تأثير على أرباح دار النشر.

لذلك قام Rajan بمقارنة ثلاثة خوارزميات للتنبؤ بحجم الطلب المتوقع على المجلات وهي خوارزميات التصنيف لأنشجار القرارات (cart) والشبكات العصبية وخوارزمية الجار الأقرب، وطبق هذه الدراسة على دار نشر روسية Burda، حيث قام بتحليل البيانات التاريخية لدار النشر من الشهر الأول لعام 2009 لغاية شهر كانون الأول عام 2012 حيث تحتوي قاعدة البيانات على 315 مجلة و 250 زبون للمجلات بواسطة برنامج Matlab، وقد توصل إلى نموذج للتنبؤ وكانت أفضل النتائج هي باستخدام خوارزمية الشبكات العصبية ومتوسط خطأ 6.95%.

- دراسة (Kinski A., 2016) بعنوان:

"Google Trends as Complementary Tool for New Car Sales Forecasting: A Cross-Country Comparison along the Customer Journey"

"اتجاهات غوغل كأداة تكميلية لتوقعات مبيعات السيارات الجديدة: مقارنة بين البلدان على امتداد رحلة العميل"

Master thesis, university of twente, berlin.

قدم Kinski طريقة جديدة في التنبؤ بمبيعات السيارات الجديدة في المانيا والولايات المتحدة من خلال أدوات جوجل (Google Trends) حيث قام بحساب فارق المدة الزمنية التي يستغرقها المشتري من بداية بحثه على الانترنت على السيارات للحصول على المعلومات عنها إلى القرار النهائي لشراء السيارة.

وقد تم استخدام نماذج الانحدار الخطي من أجل عملية التنبؤ وقد توصل إلى وجود علاقة كبيرة وایجابية بين عمليات البحث على الإنترت لنماذج السيارات وبيانات مبيعات السيارات في كل البلدين بدقة تصل إلى 68.5% وقد أدى تطبيق الفارق الزمني إلى تحسين صحة ودقة نماذج التنبؤ التي تشمل بيانات الإنترت، وتشدد الرسالة على قيمة وضرورة ضبط بيانات استعلام البحث للتنبؤ بالمتغيرات الاقتصادية، ولكنها تشير وعي الباحثين والعملاء بعدم الاعتماد بشكل أعمى على بيانات الإنترت، وتشير النتائج إلى أن السعر يتأثر بالثقافة الوطنية.

- دراسة (Carty D., 2011) بعنوان:

"An Analysis of Boosted Regression Trees to Predict the Strength Properties of Wood Composites"

"تحليل أشجار القرار المعززة للتنبؤ بخواص قوة مركبات الخشب"

master thesis, University of Tennessee, Knoxville.

تشكل صناعة المنتجات الخشبية مساهمة كبيرة في الاقتصاد الأمريكي حيث ساهمت بنسبة 6% من إجمالي الناتج المحلي الإجمالي للصناعات التحويلية في الولايات المتحدة الأمريكية ومن أجل المحافظة على القدرة التنافسية لهذه الصناعة يجب تخفيض التكاليف والمحافظة على جودة المنتج، حيث أن التحدي الذي يواجه هذه الصناعة هو المعرفة الجيدة بعوامل الطبيعة المؤثرة على الاخشاب وبين خصائص جودة المنتج النهائي حيث يهدف هذا البحث إلى التنبؤ بخصائص قوة المنتج النهائي وتم دراسة معامل المتانة ومعامل التمزق، أجريت الدراسة على أحد مصانع الخشب في الولايات المتحدة الأمريكية (particleboard) واستخدمت في التنبؤ أشجار القرارات وقد تم اخذ معياريين لتقييم النتائج هما الجذر التربيع لمتوسط الخطأ والنسبة المئوية لمتوسط الخطأ للعاملين السابقين، وقد تراوحت قيمته بين 0.99 إلى 1.443 ميغا بكسيل ونسبة الخطأ تتراوح بين 7.9% و 11.6% بينما معامل التمزق من 0.074 و 0.108 و تراوحت النسبة بين 12.7% و 18.6%.

ما يميز هذا البحث عن الدراسات السابقة:

هو دعم عملية الترويج في الهايبر ماركت وذلك من خلال دراسة تحليل بيانات سلة التسوق للهايبر ماركت باستخدام قواعد الاقتران بالإضافة إلى بناء نموذج للتنبؤ بحجم المبيعات الهايبر ماركت بالاعتماد على عدد من المتغيرات بشكل يناسب عملية التنبؤ والاستفادة من بيانات تحليل قواعد الاقتران في بناء نموذج التنبؤ، إضافة أنه لا يوجد أي دراسات في سوريا استخدمت هذه التقنيات للتوصل إلى نتائج في الهايبر ماركت.

الفصل الأول

التَّنْبُؤُ بِالْمَبْيَعَاتِ حَارِفٌ لِلْمَهَامَاتِ

Sales Forecast

في الوقت الراهن أصبحت المعلومات المتعلقة بالسوق على جانب عظيم من الأهمية في إدارة وتحطيم المشروعات المختلفة، وعملية التنبؤ بالمبيعات والطلب المستقبلي هي الداعمة الأساسية لتحطيم المشروعات، كما أنها ضرورية من أجل تقدير متطلبات السعة الإنتاجية، وكذلك توفير التنبؤ بالتدفقات النقدية، لذلك يجب أن تحظى بأهمية قصوى أيًّا كان حجم المشروع ونوعية نشاطه، فلا يجب تحت دعوى أن المشروع صغير وناشئ أن يتم الاستهانة بهذه الخطوة الهامة.

فعندما تكون قد حصلت على كل المعلومات حول خطة التسويق الخاصة بك، يمكنك البدء في صياغة خطة المبيعات لمُنتجاتك أو خدماتك خلال الأشهر الائتي عشر الأولى لمشروعك - أي السنة الأولى - مما يمثل نقطة البدء لتنظيم المشروع من الناحية المالية، لذلك يجب أن تعمل جاهدًا على جمع المعلومات الخاصة والحقائق والبيانات المتعلقة بالخطة البيعية والقيام بخطوات التحليل لها، حتى تسترشد بها عند تقرير سياسات المشروع الناشئ المستقبلي.

ستتناول في هذا الفصل الحديث عن التنبؤ من خلال النقاط التالية:

المبحث الأول: التنبؤ بالمبيعات وطرقه.

المبحث الثاني: الموازنة التقديرية للمبيعات.

المبحث الأول

التنبؤ بالمبيعات وطرقه

Sales forecast and methods

: (Introduction) 1.1.1 مقدمة

لقد احتاج الإنسان على الدوام إلى التنبؤ بأحداث المستقبل وذلك لأن خبرته بالماضي أكدت له على أن المعرفة المبكرة بالمستقبل يمكن أن توفر له فرصة أفضل للاستعداد لهذه الأحداث المستقبلية، لهذا نجد أن المجتمعات في الماضي حافلة بالذين يقومون بقراءة الكف وأعمال التنجيم وغير ذلك الكثير من الأساليب التي تعتبر اليوم بدون أساس علمي إلا أنها كانت تنبئ حاجة مهمة من حاجات الإنسان في الاطلاع على المستقبل والاستعداد له، ولكن مع التطور وتقدم خبرة الإنسان واستخدامه للأساليب الكمية المقدمة أصبح التنبؤ أداة فعالة أكثر علمية ودقة في توقع الأحداث المستقبلية مما ساعد على زيادة استعداد الأفراد وكذلك الشركات للتغيرات المتوقعة في المجالات المختلفة ومنها التغيرات في السوق وحجم ونمط الطلب على المنتجات.

ومن المعلوم أن السبب الأساسي لوجود أي منظمة هو تقديمها لسلعة أو خدمة للمجتمع، وباعتبار أن أي نشاط اقتصادي يقوم بالأساس على تلبية احتياجات ورغبات المستهلكين كماً ونوعاً من خلال إنتاج السلع أو الخدمات بالكميات والمواصفات المطلوبة، فإن الإشكالية هي أن الكمية التي يحتاجها المستهلكون تعتبر من الأمور المستقبلية لكن لابد من معرفتها أو على الأقل تقديرها حتى تستطيع المنظمة أن تعمل وتنتج لتلبية هذه الاحتياجات، فالطلب على المنتجات عادة ما يكون عرضة للزيادة أو النقصان تبعاً لعوامل وظروف مختلفة، لذلك فإن مدى قدرة إدارة المنظمة على التنبؤ بمستوى هذا الطلب سيكون له آثار هامة على مستقبلها بشكل عام وسيمثل الأساس الذي ستبني عليه خططها وبرامجها وقراراتها.

ويعتبر التنبؤ بالطلب (تقدير المبيعات) أهم وظيفة في الهرم الوظيفي هذا ما أكدته "فابيل" ¹ في قوله «أن يأتي التنبؤ في مقدمة كل ما يجب أن تقوم به إدارة المؤسسة وهو ضروري للشركات المختلفة» ويمكن أن نميز حاجة هذه الشركات للتنبؤ في حالتين: ²

¹ MAKRIDAKIS S., 1994 - FORECASTING : ITS ROLE AND VALUE FOR PLANNING AND STRATEGY, INSEAD, Fontainebleau, France.

² المتنى حسان..، 2009 - التنبؤ وتطبيقاته في إدارة الأعمال، رسالة ماجستير، جامعة دمشق.

أولاً- الشركات في طور الانشاء: رغم أن الشركات لا تمتلك بيانات تاريخية عن حجم الطلب ونمطه في الفترة الماضية إلا أنها تقوم بالتنبؤ الذي على أساسه يتم تحديد حجم المصنع ونطء الإنتاج والتنظيم الداخلي وغيرها، لهذا نجد أن هذه الشركات تستعين بأساليب عديدة لتحقيق دقة أكبر في التنبؤ مثل القيام ببحوث السوق، والاستفادة من البيانات التاريخية للشركات المشابهة، دراسة تجارب الشركات التي دخلت السوق بمنتجات لم يكن لها نظير في السابق، الاستعانة بخبرة رجال المبيعات والموزعين في سوق المنتوج، والدراسة التحليلية للظروف الاقتصادية والاجتماعية كالسكن، الاستثمارات، مستوى المعيشة في البلد.

ثانياً - الشركات في طور التشغيل والانتاج: تملك هذه الشركات قاعدة بيانات عن الفترة الماضية تكون أساساً جيداً لدقة التنبؤ عن الأحداث المتوقعة في المستقبل. ورغم أن معرفة الماضي ليست كافية لمعرفة المستقبل، إلا أن هذا لا يلغى أهمية التنبؤ، لأن الشركات الحديثة بدون التنبؤ سوف تتعامل مع المجهول وعدم التأكد المطلق. إن الخبرة المتراكمة الواسعة في مجال التنبؤ والتطور في أساليبه جعلت من الممكن تحسين درجة الدقة في هذه التنبؤات بالاستناد إلى خبرة وبيانات الفترة الماضية.

2.1.1 مفهوم التنبؤ (Concept of Forecasting)

هناك العديد من التعريف للتنبؤ نذكر منها:

- هو تقدير الحوادث المستقبلية بهدف استخدامه في أغراض التخطيط، فهو تقدير أو تخمين لما سيحدث في المستقبل وهو يسبق التخطيط، وهو يعتمد على استقراء المستقبل.¹
- هو محاولة عقلانية لتقدير التغيرات المستقبلية المحتملة من خلال معرفة المتغيرات السلوكية وغير السلوكية لتلك الظاهرة.²
- نشاط ذهني مرتبط بوجود الإنسان، وهو نتيجة لارتباط النشاط الإنساني بعنصر الوقت، وهو التوقع للتغيرات التي تحدث مستقبلاً، تؤثر بشكل مباشر وغير مباشر على النشاط.³

¹ دودين أحمد يوسف، 2012 - إدارة الإنتاج والعمليات، عمان الأردن، دار الأكاديميون للنشر والتوزيع، ص 92.

² كاظم أموري هادي، 2009 - مقدمة في القياس الاقتصادي، زهران للنشر، عمان، الأردن، ص 86.

³ الهواسي محمود، البرزنجي حيدر، 2014 - مبادئ علم الإدارة الحديثة، دار الكتب والوثائق العراقية، بغداد، العراق، ص 60.

- هو فن وعلم التوقع بالأحداث المستقبلية، هو فن لأن الخبرة والحس والتقدير الإداري له دور في التنبؤ وفي اختيار الأسلوب الملائم في التنبؤ، وهو علم لأنّه يستخدم الأساليب والطرق الموضوعية الرياضية والإحصائية في التنبؤ مما يرفع من درجة الدقة ويقلص من التحيز.¹

وتعتبر عملية التنبؤ بالمبيعات من بين أهم التطبيقات التي يستخدم فيها التنبؤ.

3.1.1 تعريف التنبؤ بالمبيعات (Definition of sales forecast):

التنبؤ هو أهم الوسائل التي تمكن المؤسسة من إعطاء نظرة مستقبلية لما ستكون عليه نشاطاتها بناء على معلومات حول الماضي والحاضر وكذلك العوامل المؤثرة في هذه المتغيرات، وهناك أكثر من تعريف للتنبؤ بالمبيعات:²

- التنبؤ بالمبيعات هو تقدير لكمية أو قيمة المبيعات المتوقعة في المستقبل، والتي يمكن أن تتم في ظل الظروف الاقتصادية والاجتماعية المحتملة.
- التنبؤ بالمبيعات هو نقطة الانطلاق نحو تقرير نشاط المشروع، من إنتاج وتسويق وتمويل وإعداد ميزانية التقدير، ومختلف برامج الإنتاج والمخزون.

أيضاً يرى الأستاذ هنري فايلو الذي يعتبر الأب الحقيقي لعلم الإدارة أن قوة التنبؤ بالأشياء قبل حدوثها هو جوهر الإدارة. فالقيام بعملية التنبؤ بشكل صحيح يوصلنا إلى الهدف المطلوب إذا قمنا فعلاً باتباع الطرق السليمة والصحيحة واعتمدنا على الأسس العملية والطرق الرياضية والإحصائية لعملية التنبؤ. إن التفكير المنظم السابق لعملية التنفيذ والاستعداد والتهيؤ للمستقبل قائم على التنبؤ والاحتمالات التي تضعها الجهات الإدارية على ضوء الإمكانيات المتاحة. إن الغرض الأساسي من جراء عملية التنبؤ بالمبيعات هو الوصول إلى رقم المبيعات التقديرية سواء كان هذا الرقم يعبر عن كمية أو قيمة لفترة زمنية قادمة. ويعطي التنبؤ بالمبيعات مؤشراً عن حجم المبيعات المتوقعة الذي يمكن تحقيقه من سلعة أو مجموعة سلع خلال فترة زمنية مقبلة في ضوء خطة تسويقية معينة.³

¹ نجم عبود نجم، 2013 - مدخل إلى إدارة العمليات، دار المناهج للنشر والتوزيع، عمان، ص 158-159.

² كاظم اموري هادي، 2009 - مقدمة في القياس الاقتصادي، مرجع سبق ذكره، ص 86-88.

³ عبيدات محمد، الضمور هاني، حداد شفيق - 2013 ، إدارة المبيعات والبيع الشخصي، دار وائل للنشر، عمانالأردن، ص 180.

ومن وجهة نظر الباحث فإن التنبؤ بالمبيعات هو تقدير لكمية أو قيمة المبيعات من منتج معين أو مجموعة من المنتجات خلال فترة زمنية محددة في المستقبل، وذلك بالاعتماد على بيانات ومعلومات سابقة وظروف مشابهة باستخدام تقنيات تنبؤ مختلفة.

4.1.1 أهداف التنبؤ بالمبيعات (Sales Prediction Objectives):

تسعى المؤسسة من خلال عمليات التنبؤ من الوصول إلى رقم تقديرى للمبيعات، ومن خلال هذا الرقم من الممكن أن تتحقق العديد من الأهداف، وفيما يلى بعض الأهداف التي من الممكن أن تتحققها المؤسسات:¹

1. يعتبر التنبؤ بالمبيعات الأساس الذي يعتمد عليه في عمليات التخطيط في كافة الأنشطة التي تمارسها أقسام وإدارات المنشأة، حيث على ضوء هذا التنبؤ يتم صياغة خطط الإنتاج والمشتريات والتسويق والتمويل.
2. تستطيع المنشأة من خلال عملية التنبؤ بالمبيعات تقدير تكاليف الأنشطة التي ستقوم بتنفيذها، وبالتالي تتمكن من تحديد مصادر الحصول على الأموال ويتم اعداد الموازنة المالية.
3. تستطيع المنشأة من خلال عملية التنبؤ تحديد الأرباح المتوقعة من المبيعات في نهاية الفترة الزمنية التي تعطيها عملية التنبؤ، وذلك بطرح تكاليف المبيعات المتوقعة من الإيرادات المتوقعة من المبيعات.
4. يساعد التنبؤ بالمبيعات الإدارية في مراقبة نشاط إدارة المبيعات ورجال البيع وتحديد مدى كفاءتهم في تنفيذ المهام المسندة إليهم وذلك لأن التنبؤ يوفر الأساس الذي يستخدم في مقارنة المبيعات المحققة من المبيعات المتوقعة.
5. يساهم التنبؤ في توجيه جهود الأفراد العاملين وتوظيفها لخدمة تحقيق الأهداف وترشيد قرارات الادارة المتعلقة الإنفاق على مختلف الأنشطة.

5.1.1 أهمية التنبؤ بالمبيعات وعلاقته بخطيط المبيعات (Importance of sales forecasting and its relation to sales planning)

إن التنبؤ بالمبيعات يعد عملية أساسية ولها أولوية على جميع التنبؤات الأخرى، فالتنبؤ بالمبيعات طويلة الأجل يساعد على تخطيط النفقات الاستثمارية (الاستثمار الرأسمالي)، والتنبؤ بالمبيعات في

¹ عبد الحميد طلعت أسد.، 2010 - التسويق الفعال ، رؤية للنشر والتوزيع، القاهرة.

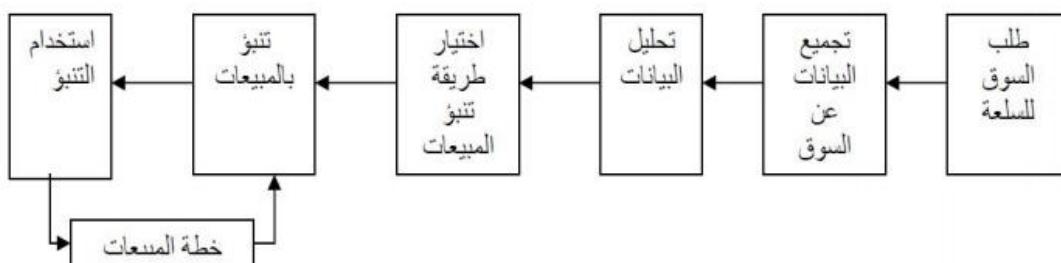
² موسى غانم فنجان.، 1999 - إدارة المبيعات والاعلان، دار الحكمة للطباعة، بغداد، العراق، ص 90.

الأجل القصير يساعد على تحديد الكميات المطلوب إنتاجها، ومن ثم تحديد المستلزمات السلعية المطلوب توافرها، والقوى العاملة المطلوب وجودها.

إذ نستطيع القول أنه هناك علاقة بين التنبؤ بالمبيعات وتحطيط المبيعات، إذ يعتمد تحطيط المبيعات على النتائج الحاصلين عليها من عملية التنبؤ، والتنبؤ بالمبيعات يعتمد هو أيضاً على خطط المبيعات التي تم إعدادها في السابق، باعتبار أن تلك الخطط تمثل أحد مصادر المعلومات التي تقوم الجهات الموكلا لها مسؤولية التنبؤ، حيث تقوم بدراستها وتحليلها واعتمادها كمؤشر هام في التنبؤ بالمبيعات في المستقبل.

ويعرض الشكل رقم (1) العلاقة المتبادلة بين التنبؤ بالمبيعات وتحطيط المبيعات، حيث يظهر منه أن نقطة البداية في عمليات التنبؤ بالمبيعات وتحطيطها تكون بدراسة السوق الذي تسوق فيه المنشأة سلعها.

ويطلب هذا تجميع البيانات عن السوق وعن المتغيرات المؤثرة في الطلب على المبيعات، وبعدها يتم تحليل البيانات بهدف الحصول منها على مؤشرات تستخدم في عملية التنبؤ، وتقوم الجهة المكلفة بالتنبؤ باختيار إحدى طرق التنبؤ انطلاقاً من المؤشرات المتحصل عليها، وبعد الانتهاء من إعداد التنبؤ بالمبيعات يتم إعداد خطط المبيعات على ضوء نتائج عملية التنبؤ.



الشكل رقم (1) العلاقة بين التنبؤ بالمبيعات وتحطيط المبيعات

مصدر: موسى غانم، 1999 - إدارة المبيعات والإعلان، دار الحكمة للطباعة، بغداد، العراق، ص 90.

6.1.1 أبعاد عملية التنبؤ بالمبيعات (Dimensions of Sales Forecasting)¹:

تنقسم عملية التنبؤ بالمبيعات إلى ثلاثة أبعاد رئيسية هي: التنبؤ الكلي بالنشاط الاقتصادي العام والتنبؤ بحجم المبيعات على مستوى القطاع والتنبؤ بحجم المبيعات على مستوى المنشأة.

1. على مستوى الاقتصاد العام:

نظراً للأهمية المتزايدة للتغيرات الاقتصادية على نتائج أعمال المشروع فإن البداية السليمة لإعداد تنبؤ جيد بالمبيعات يجب أن تتمثل في التنبؤ بالمناخ الاقتصادي العام وما يرتبط بذلك من تحليل للسياسات المالية والنقدية والإتفاق الاستهلاكي والاستثمارات وغير ذلك من المؤشرات الأساسية في تحديد اتجاه النشاط الاقتصادي للدولة. ومن الأهمية بمكان في هذا الصدد أن يختار من يقوم بالتنبؤ المؤشرات ذات التأثير على طبيعة نشاطه وعلى سبيل المثال بالنسبة لمشروع يقوم بإنتاج السيارات فإنه من الضروري أن يميز عند التنبؤ بمبيعاته بين عنصرين هما:

- المبيعات من الإطارات التي مستخدمة في السيارات الجديدة وهي تتأثر بالحالة الاقتصادية العامة لأن هذه الحالة تؤثر دون شك على مبيعات السيارات الجديدة.
- المبيعات بفرض الاحلال وهذه تتأثر بعدد السيارات التي بيعت في الفترة القصيرة السابقة.

2. على مستوى القطاع:

يتطلب الأمر من المشروع أن يقوم بتقدير الطلب الكلي للسوق أي مبيعات الصناعة التي ينتمي إليها وذلك حتى يمكنه من تحديد الفرص البيعية المتاحة له، ويمكن تعريف الطلب الكلي بالنسبة لمنتج معين بأنه الحجم الكلي الذي سيشتري من هذا المنتج بواسطة مجموعة محددة من المستهلكين في منطقة جغرافية محددة وفي فترة زمنية محددة وفي مناخ تسويقي وفي ظل برنامج تسويقي محدد ويعتبر تقدير الطلب الكلي للصناعة من أهم العقبات التي تواجه من يقوم بالتنبؤ بمبيعات المشروع خاصة وأن هناك صعوبات كبيرة في تحديد هذا الطلب.

3. على مستوى المنشأة:

يعتبر التنبؤ بالمبيعات على مستوى المنشأة بأصنافها المختلفة ومناطقها المتعددة من الأمور الازمة لإعداد الميزانيات التقديرية على مستوى المنشأة ككل، وقد تناولنا في بحثنا هذا التنبؤ على مستوى المنشأة من خلال التنبؤ بمبيعات الهايبر ماركت.

¹رابعة علي - ذياب فتحي، 1990 – إدارة المبيعات، منشورات مركز الخوارزمي . عمان.

7.1.1 العوامل المؤثرة في دقة التنبؤ بالمبيعات (Factors affecting the accuracy of sales forecasting)

تسعى إدارة المبيعات في كل المنشآت لأن تصل إلى تنبؤ دقيق عن المبيعات ذلك أن أي خطأ في المبيعات له تأثير على كل النشاطات الأخرى في المنشأة وليس على إدارة المبيعات فقط، لهذا نجدها لا توفر جهداً في استخدام كل الأساليب الكمية وغير الكمية، وتوظف كل الخبرات والمؤهلات التي تجعلها قادرة على تحقيق هذا الهدف.

وتشير الدراسات¹ التي تناولت التنبؤ بالمبيعات أن الأخطاء الكبيرة في التنبؤ بالمبيعات سواء بالزيادة أو النقصان يكلف المنشأة الكثير من الأضرار الاقتصادية والمعنوية. ويمكن تلخيص النتائج والأضرار التي تحصل عندما تكون المبيعات المتوقعة أكبر من المبيعات المحققة فيما يلي:

- تجميد مبالغ في المخزون السلعي.
- تزايد احتمال تعرض المخزون السلعي للتلف والتقادم.
- ارتفاع المبالغ المدفوعة كتأمين على المخزون.
- تشغيل موارد في مختلف التخصصات أكثر من المطلوب.
- ارتفاع تكاليف التخزين.

أما النتائج التي تحصل في حالة التقديرات المتوقعة للمبيعات أقل من المحققة فهي:

- فقدان سمعة المنشأة في الأسواق والأوساط التجارية مستقبلاً نتيجة لعدم إرضاء العملاء والمستهلكين.
- فقدان فرص الحصول على الأرباح المتوقعة من المبيعات.
- تضطر المنشأة في بعض الحالات بسبب رغبتها في تلبية حاجة العملاء إلى شراء المواد والبضائع بأسعار مرتفعة وتشغيل العمال أوقاتاً إضافية، وبالتالي دفع أجور إضافية - (تكلفة زائدة).
- تحدث الزيادة الحقيقة في الطلب إياكاً في جميع الأنشطة.

¹ عبد الحميد طلعت اسعد، 2000 - دليل مدير المبيعات الفعال : كيف تدير العملية البيعية بكفاءة، مؤسسة الأهرام، مصر.

- تؤدي الزيادة في الطلب إلى الإسراع في توفير البضاعة للعملاء مما قد يؤدي إلى عدم قدرة المنشأة على توفيرها بالنوعية المطلوبة، الأمر الذي يؤدي في المدى البعيد إلى خسارة العملاء.

يتأثر التنبؤ بالمبيعات بنوعين من العوامل هي خارجية وداخلية¹.

1.7.1.1 العوامل الخارجية : (Facteurs externes)

وتشمل جميع العوامل التي تؤثر في المبيعات وليس للمنشأة سيطرة عليها، ونذكر منها ما يلي:

- العوامل الاقتصادية: مثل مستوى الدخل القومي، دخل الفرد، القوة الشرائية، حجم الاستثمار وتوزيعاتها في القطاعات الاقتصادية، خطط الدولة فيما يخص الاستيراد والتصدير، الضرائب، تقلبات الأسعار، المنافسة، وخطط الدولة واتجاهاتها بخصوص التدخل في الحياة الاقتصادية وبرامجهما التي تتعلق بدعم القطاع الخاص والقطاع المختلط في الحياة الاقتصادية وبرامجهما التي تتعلق بدعم القطاع الخاص والقطاع المختلط والقطاع التعاوني وكذلك مستوى الاستخدام وتوفير فرص العمل للقادرین على العمل والراغبين فيه.
- العوامل الديموغرافية: والتي تهتم بدراسة السكان ونسب النمو في السكان، وتوزيع السكان جغرافياً، وحسب الجنس، ودرجة تأثير كل هذه العوامل على طلب السلع في المستقبل.
- العوامل الاجتماعية: وتشمل العادات والتقاليد.
- العوامل الثقافية والعلمية والتقنية: وتتضمن جميع العوامل التي لها علاقة بالمستوى الثقافي والعلمي والتقني السائد في المجتمع، حيث يساعد هذا في تحديد نوعية السلع المطلوبة من أفراد المجتمع.
- العوامل الطبيعية: وتشمل جميع العوامل التي لها علاقة بالمناخ والتضاريس، والتي يكون لها تأثير في الطلب على السلع، حيث يحدد المناخ نوع السلع المطلوبة من الأفراد ومقدار الطلب والأوقات التي يزداد فيها الطلب فلا يمكن تسويق مكيفات الهواء في المناطق الباردة مثل دول شمال أوروبا حيث يكثر الطلب على المدافئ.

2.7.1.1 العوامل الداخلية : (Facteurs internes)

وتشمل جميع العوامل التي تتعلق بالإمكانيات المادية والبشرية المتاحة للمنشأة، والتي تؤثر في تحديد قدرتها في مواجهة العوامل الخارجية وطرح السلع المطلوبة من الأفراد، وهذه العوامل هي:

¹ رياض قادری، 2011 - طرق وأساليب التنبؤ عن المبيعات دراسة حالة، رسالة ماجستير كلية علوم التسيير والاقتصاد، الجزائر، ص49.

- الإمكانيات المالية للمنشأة.
- كفاءة العاملين في المنشأة بصفة عامة، وكفاءة القوى التجارية وكفاءة الأجهزة التي تتولى التنبؤ بالمبيعات وتحطيم المبيعات وبحوث السوق.
- قدرة المنشأة على طرح سلعة جديدة أو تطوير السلع الحالية بهدف مواجهة المنافسة.
- سياسات المنشأة المتعلقة بالأرباح التي تمنحها لل وكلاء والموزعين.
- سياسات التسويق والاعلان في المنشأة.
- طرق ومنافذ التوزيع التي تعتمد عليها المنشأة في توصيل البضائع إلى المستهلكين والعملاء حالياً وفي المستقبل.
- تكاليف التسويق.

8.1.1: (Duration of prediction)

تختلف طول مدة التنبؤ من قطاع اقتصادي لأخر ومن صناعة لأخرى ومن منشأة لأخرى، نظراً لاختلاف الصناعات من حيث الظروف التي تعمل ضمنها أو السلعة التي تقوم بإنتاجها أو تسويقها لهذا نجد ضمن بعض المنشآت تكون مدة التنبؤ بالمبيعات فيها تغطي سنة واحدة وأخرى تزيد فيها هذه المدة عن السنة وقد تقل عن السنة في منشآت أخرى.

فمثلاً المنشآت التي تسوق السلع الرأسمالية كالعتاد يكون التنبؤ فيها يغطي خمسة سنوات فما فوق، أما المنشأة التي تسوق السلع الموسمية كالدافئ والمكيفات، فلا تزيد فترة التنبؤ فيها عن موسم واحد، لكن لابد من التأكيد على أن أغلب المنشآت تعتمد التنبؤ لمدة سنة، وينتج مما تقدم أن هناك نوعين من التنبؤ بالمبيعات وفقاً لمعايير المدة¹، وهما:

- **التنبؤ قصير الأجل:** ومدته سنة واحدة أو أقل، وهو الذي يعتمد عليه أساساً في إعداد موازنة التشغيل وفي تحديد التكاليف المعيارية للأسعار الجارية والتدفق النقدي.
- **التنبؤ طويل الأجل:** هو التنبؤ الذي يغطي مدة تزيد عن سنة، وهذا النوع هو الذي يعتمد عليه أساساً في تحطيم الاحتياجات من المعدات والمباني المطلوبة في العملية الإنتاجية في المستقبل البعيد.

¹ مخرش عبلة، 2006 – تدبر نموذج للتنبؤ بالمبيعات باستخدام السلسل الزمنية (نماذج بوكس وجينكينز) دراسة حالة الشركة الوطنية للكهرباء والغاز، رسالة ماجستير، كلية الحقوق والعلوم الاقتصادية، الجزائر.

9.1.1 خطوات التنبؤ : (Steps to Forecasting)

تم عملية التنبؤ وفق خطوات محددة يمكن عرضها في الآتي:¹

1. تحديد الغرض من القيام بالتنبؤ، وذلك لأن المعلومات الخاصة بالتنبؤ يستخدمها مدراء الوظائف المختلفة في مباشرتهم لوظائفهم، واتخاذهم لقراراتهم الإدارية، فمثلاً التنبؤ بالإيراد السنوي قد يكون مفيد لمدير الإنتاج والعمليات لاتخاذ القرارات المتعلقة بتحديد مستويات الإنتاج والاحتياجات من العمالة، لكن مدير الإنتاج والعمليات قد لا يستفيد من الرقم الإجمالي للتنبؤ مثل مدير التسويق، ويحتاج إلى معلومات أكثر تفصيلاً ل يستطيع إعداد جداول الإنتاج التفصيلية بما يتفق مع احتياجات المستهلكين.
2. جمع البيانات التاريخية سواء عن الاتجاهات الاقتصادية من المستندات الحكومية أو سجلات الشركة، وفي حالة المنتجات الجديدة والتي لا تتوفر عنها البيانات الإحصائية التاريخية قد يكون من الضروري استخدام البيانات المتاحة عن منتجات مشابهة أو منافسة.
3. عرض البيانات التاريخية على رسم بياني لتحديد مدى وجود نمط معين لاتجاه البيانات، سواء أظهرت وجود دورة معينة للبيانات أو وجود بيانات باتجاهات موسمية تمكن من توقع البيانات في المستقبل.
4. اختيار نموذج للتنبؤ والذي قد يستخدم في المواقف الإدارية المختلفة وعلى الإدارة تطبيق النموذج الذي يتماشى مع احتياجاتها.
5. إجراء التجارب التي تظهر مدى صحة الطرق التي استخدمت في التنبؤ بالقيم الحقيقة التي ظهرت خلال الفترة الماضية.
6. متابعة نتائج تطبيق أسلوب التنبؤ عن طريق تسجيل الأداء الفعلي ومراقبة خطأ التنبؤ، وعلى المدير أن يقرر على فترات ما إذا كانت عمليات التنبؤ الحالية تؤدي إلى تنبؤ مقبول، فإذا كان الأمر كذلك فإن الأسلوب الذي تم اختياره يستمر في تطبيقه، أما في حالة عدم قبول مقدار الخطأ لتجاوزه ما هو مسموح به في هذه الحالة نحتاج إلى أسلوب جديد وهنا نعود إلى الخطوة الرابعة، وهكذا تتكرر الدورة في كل مرة.

¹البكري صونيا، 2001 - إدارة العمليات والإنتاج، الدار الجامعية، الإسكندرية، مصر.

10.1.1 طرق التنبؤ بالمبيعات : (Methods of sales forecasting)

تتمثل هذه الطرق بنوعين هما الطرق النوعية والطرق الكمية والطرق التقنية.

1.10.1.1.1 الطرق النوعية (الوصفية) :

التي تعتمد على الخبرة ورأي الأفراد داخل وخارج المؤسسة وحسب المستوى الهرمي للقرارات ومنها نجد:

1.10.1.1.1.1 الرأي الجماعي¹:

وفي هذه الطريقة يلتقي مجموعة من الأشخاص من ذوي الخبرة والخلفية العملية، ويطلب من كل منهم وضع تقديرات لقيمة المبيعات المتوقعة، وبعدها يتم التوصل إلى نهاية متوقعة للمبيعات بإحدى الطريقتين:

- أخذ متوسط للتقديرات الفردية.
- قيام رئيس المجموعة بمراجعة جميع التقديرات الفردية لأعضاء الجماعة ثم تقدير قيمة المبيعات المتوقعة بناء على خبرته الشخصية.

إن هذه الطريقة مفيدة في التوصل إلى تقدير المبيعات في وقت قصير، كما تتميز بانخفاض تكلفة القيام بالتتبؤ، وتتمي الروح المعنوية لدى أعضاء الجماعة ولكن ما يؤخذ على هذه الطريقة صعوبة تحديد المسؤول عن أخطاء التقدير، وكذلك وجود عنصر التحيز الشخصي في عملية تقدير المبيعات.

2.1.10.1.1.2 أسلوب لجنة الخبراء² : The Expert Panel

وهنا يجتمع مجموعة خبراء (8-12 خبيراً) في مكان واحد لتقدير حجم الطلب المتوقع على السلعة أو الخدمة، ويكون لهذه المجموعة منسق يطلب من كل عضو في المجموعة تقديم فكرة رئيسية واحدة، ثم تبدأ مناقشة هذه الأفكار ويمكن حذف أو دمج بعض هذه الأفكار، بعدها يجري تصويت سري على الفكرة الرئيسية لتقدير حجم الطلب المتوقع.

ميزة هذه الطريقة هي الاستفادة من آراء مجموعة كبيرة من الخبراء واتخاذ القرار في نهاية الاجتماع أما سلبياتها فتمثل في ارتفاع التكلفة المادية نتيجة لحضور الخبراء من أماكن مختلفة.

¹ الحسين محمد ابديوي..، 2012 - تخطيط الإنتاج ومراقبته، دار المناهج للنشر والتوزيع، عمان، الأردن، ص 25.

² الكرخي مجید..، 2014 - تخطيط وتقدير البرامج، دار المناهج للنشر والتوزيع، عمان، الأردن، ص 27.

3.10.1.1 طريقة دلفي:¹

تم تطوير طريقة دلفي في عام 1964 من قبل مؤسسة البحث والتطوير الأمريكية المعروفة بمؤسسة راند، وقد استخدمت لأول مرة في التنبؤ التكنولوجي حيث أن أساس هذه الطريقة هو اشتراك عدد معين من الخبراء في عملية التنبؤ بظاهرة معينة وذلك عن طريق مراسلة تتم بالمراحل التالية:

- يتم اختيار شخص يكون مسؤولاً عن مهمة القيام بعملية التنبؤ ويسمى بالمنسق يتميز هذا الأخير بدرجة عالية من الخبرة والمعرفة بالظاهرة محل التنبؤ.
- يقوم المنسق بإرسال استفسارات إلى الخبراء في صورة قائمة أسئلة لإبداء الرأي حول الظاهرة محل التنبؤ.
- عند وصول الإجابات التحريرية من قبل الخبراء ويقوم المنسق بدراسة كل المراسلات وتبويب الإجابات ثم إرسال استفسارات لهؤلاء مع تزويدهم بالمعلومات المتتجدة والمستوحاة من قبل بعض الخبراء المشاركون في عملية التنبؤ ثم يطلب منهم إبداء الرأي حول الظاهرة مجدداً مع توضيح المبررات.
- يتم تكرار الخطوة السابقة عدة مرات حتى يتم التوصل إلى درجة كبيرة من الاتفاق في تقديرات الخبراء حول الظاهرة محل الدراسة.

من مزاياها:

- أنها تساهم في الاستفادة من آراء مجموعة كبيرة من الخبراء المختصين.
- انخفاض التكلفة المادية نتيجة تبادل الآراء عن طريق المراسلة.
- الانفراد والحيادية وعدم التأثير على النتيجة لعدم الاجتماع.

ومن عيوبها: أنها تستغرق فترة زمنية طويلة في عملية اتخاذ القرارات.

2.10.1.1 النماذج الكمية:

تعتمد الأساليب الكمية على استخدام النماذج الرياضية في تحليل المتغيرات الخاصة بالمؤسسة وهناك الكثير من الطرق الإحصائية ذكر منها:

1.2.10.1.1 طريقة المتوسط البسيط:¹ يتم حساب المتوسط الحسابي للمتغير المدروس (الظاهرة) لفترات زمنية سابقة، ثم يستخدم هذا المتوسط للتنبؤ بالفترة الزمنية اللاحقة وهو من أبسط الطرق الإحصائية.

¹ نجم عبود نجم، 2013 - مدخل إلى إدارة العمليات، مرجع سابق ذكره، ص 165.

2.2.10.1.1 طريقة المتوسطات المتحركة البسيطة²

وهي أكثر النماذج استخداماً حيث تستخدم عند التنبؤ بقيمة متغير ما لفترة زمنية، فهي عبارة عن سلسلة من الأوساط الحسابية لعدد محدود من البيانات الزمنية، وهي طريقة سهلة التطبيق لكن من عيوبه تأخذ كل المشاهدات بنفس الوزن.

3.2.10.1.1 طريقة المتوسطات المتحركة المرجحة³

تاريجية وزن معين في السلسلة الزمنية ويعتبر هذا من نفائص الطريقة السابقة حيث يترجم الوزن بمعامل الترجيح باعتبار أن السلسلة تتخللها تقلبات حادة خلال فترة زمنية محددة، أي تأخذ هذه التبدلابات بعين الاعتبار وترى على أنها الوسط الذي يتم تعديله بشكل مستمر مع مرور الفترات الزمنية عن طريق تغيير الأرقام الذي يحسب على أساسه وذلك بإضافة رقم جديد وإسقاط رقم قديم.

4.2.10.1.1 طريقة التمهيد الأسني البسيط⁴

إن الطرق السابقة تتطلب وجود بيانات خاصة بالمتغير لفترتين على الأقل لكي تم عملية التنبؤ بينما هذه الطريقة تستبعد هذا النقص وتبحث عن وجود ثلاثة بيانات فقط وهي القيمة الفعلية الأخيرة الخاصة بالظاهرة محل التنبؤ وأخر قيمة متوقعة ومعامل الترجح.

5.2.10.1.1 طريقة تحليل السلاسل الزمنية⁵

إن استخدام طريقة الانحدار الخطي البسيط غير كافية لإظهار آثار بعض المؤثرات النوعية الهامة الخارجية والتي قد لا يكون لها دور كبير في تقسيم قيمة المشاهدات، ومن ثم يجب استخدام نموذج السلاسل الزمنية لتحليل البيانات بنوعيها الثابت التي تكون فيها البيانات متوازية حول وسط معين وغير الثابت التي تكون فيه البيانات تتميز بوسط متحرك أو اتجاه عام.

تعتمد السلاسل الزمنية على النماذج التصاديفية وأبسطها السير العشوائي لاستخدام الانحدار الذاتي للوسط المتحرك المتكامل المعروف باسم ARIMA والقيام بالتنبؤ الملائم واختبار درجة دقتها، وقد

¹ نجم عبود نجم. 2013 - مدخل إلى إدارة العمليات، مرجع سبق ذكره، ص 167-169.

² حامد جمال. 2003 - أساليب التنبؤ، مجلة جسر التنمية، العدد الرابع عشر.

³ حامد جمال. 2003 - أساليب التنبؤ، مجلة جسر التنمية، العدد الرابع عشر.

⁴ الطائي فاصل عباس، 2009 - التنبؤ والتمهيد للسلاسل الزمنية باستخدام التحويلات مع التطبيق، المؤتمر العلمي الثاني للرياضيات والاحصاء والمعلوماتية، جامعة الموصل.

⁵ طعمة سعدية عبد الكريم، 2012 - استخدام تحليل السلاسل الزمنية للتنبؤ بأعداد المصايبين بالأورام الخبيثة في محافظة الانبار، مجلة جامعة الأنبار للعلوم الاقتصادية والإدارية، العدد 8.

كثر استخدامه في مجال الإدارة لاتخاذ القرارات خاصة في سنوات التسعينيات على أيدي "بوكس وجينكنز" الذين استبطا أفضل النماذج للتنبؤ منها:

- النماذج ذات المتغير الواحد (بمدخل واحد وخرج واحد).
- النماذج الدالة المحولة (عدة مدخلات وخرج واحد).
- نماذج الطوارئ.
- النماذج الدالة المحولة المتعددة للمتغيرات (عدة مدخلات وعدة مخرجات).

3.10.1.1 الطرق التقنية (Technical Methods) :

تعتمد الطرق التقنية على طرق تنبؤية حديثة تستخدم خوارزميات معقدة مختلفة، وهناك الكثير من التقنيات ذكر منها:

1.3.10.1.1 الشبكات العصبية (Neural Networks)¹:

تعتبر من الطرق الحديثة والفعالة في التنبؤ وهي عبارة عن نموذج رياضي يعمل على محكاة هيكلية ووظائف الشبك العصبية الحيوية، وكل كتلة بسيطة من الشبكة العصبية عبارة عن عصبون صنعي وهذا العصبون عبارة عن نموذج رياضي بسيط يمتاز بمجموعة من القواعد: الضري والجمع والتشييط، وتنقيل مدخلات العصبون أي كل قيمة مدخلة تضرب بوزن مستقل، والمرحلة الثانية من العصبون الصنعي تعبر عن عملية الجمع والتي من خلالها يتم جمع كل المدخلات المنقلة وانحرافاتها عبر تابع تشييط ويدعى أيضاً تابع المرور.

2.3.10.1.1 أشجار القرار (Decision Tree)²:

تعتبر شجرة القرار أداة تستخدم غالباً في التصنيف والتنبؤ، وهي أداة قوية جداً لتمثيل المعرفة، وتعرف شجرة القرار على أنها بنية هيكلية تستخدم لنقسام مجموعة كبيرة من السجلات إلى مجموعات صغيرة متتالية عن طريق تطبيق قواعد القرار البسطة وفق تتابع معين، وهي تقنية تصنيف مسطة وواسعة الانتشار وهي مخطط يشبه بناء الشجرة وتوي الشجرة ثلاثة أنواع من العقد:

- عقدة جذر Root Node: ليس لها تفرع قادم ولها تفرعات خارجة.
- عقد داخلة Internal Nodes: لكل منها تفرع قادم واحد فقط وتفرعات خارجية أو أكثر.
- عقدة ورقية Leaf أو طرفية Terminal: لكل منها تفرع قادم واحد فقط وليس لها أي تفرع خارج.

¹ الشرقاوي محمد علي، 2004 - الذكاء الصناعي والشبكات العصبية، مركز الذكاء الصناعي للحاسبات، مصر، ص 273.

² الطويل هالة، 2009 - التنبؤ في البيانات، دار شعاع، دمشق، ص 155-157.

لا يمكن القول بأنه هناك نقية من بين الطرق الخاصة بالتنبؤ بأنها فعالة إلا إذا حققت مجموعة من الشروط هي: الكلفة، الدقة، توفير البيانات اللازمة، الوقت المحدد لجمع المعلومات، توفر الإمكانيات اللازمة المادية والبشرية والمعنوية لقيام بعملية التنبؤ.

المبحث الثاني
الموازنة التقديرية للمبيعات
Estimated sales budget

1.2.1 مقدمة (introduction)

تعتبر عملية التنبؤ بالمبيعات الخطوة الأولى في إعداد الموازنة التقديرية للمبيعات والقاعدة التي تبني عليها باقي تقديرات الموازنات التقديرية. أثبتت جميع التجارب أن نجاح نظام الموازنات التقديرية يتوقف إلى حد كبير على مدى الدقة في عملية التنبؤ بالمبيعات المستقبلية. وتعتبر موازنة المبيعات التقديرية أساساً لإعداد موازنة المشتريات والإنتاج ومرشداً للتوسيع في المشروع باعتبارها المورد الرئيسي للأرباح والمقوضات النقدية التي يعتمد عليها لتمويل خطط المشروع.

إن الموازنة التقديرية ليست مجرد تنبؤ بالمستقبل لكن هي دراسة تهدف إلى التأثير في الظروف المستقبلية مع التعبير عنها بالخطط والسياسات التي يرسمها المشروع في مدة زمنية محددة وأداة لوضع خطة تتضمن تحديد هدف معين وتحديد الجهد وإجراءات العملية اللازمة لتحقيق الهدف.¹

من الضروري جداً عند عمل الموازنة إظهار الأرقام بالكميات والأسعار وليس فقط بالقيمة وذلك حتى تسهل مهمة تحليل نتائج الأداء الفعلي وفسح المجال أمام رجال التسويق بمعرفة الفرق بين المبيعات المتوقعة والمبيعات الفعلية هل هو ناتج عن فرق في حجم المبيعات أو فرق في أسعار البيع.

2.2.1 مفهوم الموازنة التقديرية (Concept of the estimated budget)

قبل التطرق إلى تعريف الموازنة التقديرية ننطرق إلى تعريف الموازنة حيث تعرف على أنها: عبارة عن تقرير يوضح المعلومات الخاصة باستخدام المؤسسة المتمثلة بأصولها ومصادر هذه الاستثمارات المتمثلة في الخصوم وحقوق الملكية.²

وهناك الكثير من التعريفات للموازنة التقديرية ذكر منها:

¹ عبدات محمد، الضمور هاني، حداد شفيق، - 2013 ، إدارة المبيعات والبيع الشخصي، مرجع سبق ذكره، ص209.

² الدرى مؤيد عبد الرحمن، حسين محمد سلامة، - 2013 - أساسيات الإدارة المالية، دار الراية للنشر والتوزيع، الأردن.

- خطة عمل تفصيلية محددة مسبقاً للأعمال المرغوب بتنفيذها وتوزع هذه الخطة على جميع المسؤولين حتى تكون مرشداً لهم في تصرفاتهم.¹
- قائمة بالأهداف العامة والفرعية عن فترة زمنية مستقبلية معبراً عنها مالياً، وهي تتخلل كل مستوى للنشاط تحقيقاً تكامل خطط الإيرادات وخطط المصروفات والاحتياجات من الأصول واحتياجات التمويل.²
- تعبير كمي أو مالي لأهداف المؤسسة المسطرة، كما يمكن اعتبارها خطة تفصيلية محددة مسبقاً للأعمال المرغوب في تنفيذها في المؤسسة خلال فترة زمنية معينة في المستقبل مترجم عنها في شكل نقدي.³
- تعبير كمي لخطة الأعمال وتساعد على تحقيق التنسيق والرقابة.⁴
- ويرأى الباحث يمكن تعريف الموازنة التقديرية بأنها: خطة عمل تفصيلية تقوم المنشأة بإعدادها ترجم فيها أهداف المنشأة على شكل مبالغ مالية، وهي نتائج محتملة من أجل تحقيق أفضل استخدام للموارد المتاحة خلال فترة معينة.

3.2.1 أهمية الموازنة التقديرية للمبيعات : (Importance of Estimated Sales Budget)

تعتبر الميزانية التقديرية مهمة للعديد من الأسباب منها:⁵

- 1- تعتبر نقطة الأساس في إعداد الموازنات الأخرى في المشروع.
- 2- تحقيق الأهداف البيعية بأقل التكاليف والحصول على أكبر قدر ممكن من الأرباح.
- 3- تنمية المبيعات عن طريق تحفيز مندوبي البيع لبلوغ الأهداف المحددة فيها.
- 4- التنسيق بين جهود مندوبى البيع والمناطق البيعية.
- 5- وضع الخطة الشاملة للمشروع حيث تعتبر أساساً لإعداد خطط الإدارات الأخرى.
- 6- أداة للرقابة على تنفيذ الخطط ومعرفة نقاط الضعف والعمل على علاجها في الوقت الملائم.
- 7- تقييم جهود مندوبى البيع وتحديد مسؤولية مديرى الفروع والمناطق البيعية عن النتائج التي توصلوا إليها.

¹ مصطفى يوسف كافي..، 2012 - المحاسبة الإدارية، مكتبة المجتمع العربي للنشر والتوزيع، الأردن.

² أبو زيد كمال خليفة ، مرعي عطية.، 2004 - مبادئ المحاسبة الإدارية الحديثة، الدار الجامعية، الإسكندرية، مصر.

³ عدون ناصر ..، 2004 - مراقبة التسبيير في المؤسسة الاقتصادية، دار المحمدية العامة، الجزائر.

⁴ فركوس محمد.، 2001 - الموازنة التقديرية أداة فعالة للتسبيير، ديوان المطبوعات الجامعية، الجزائر.

⁵ عبيدات محمد.، الضمور هاني.، حداد شفيق.، - 2013 ، إدارة المبيعات والبيع الشخصي، مرجع سبق ذكره، ص209-211.

4.2.1 مراحل إعداد الموازنة التقديرية للمبيعات (The stages of preparing the estimated budget for sales)

تتضمن مراحل إعداد الموازنة التقديرية للمبيعات ما يلي:

1. **تكوين لجنة الميزانية:** يتم تكوين لجنة الموازنة للإشراف على إعداد وتنسيق تقديرات الموازنة، وعادة ما تعد الميزانية لمدة سنة أو أقل، غالباً تكون هذه اللجنة برئاسة مدير عام للشركة وعضوية مدير الإنتاج ومدير المبيعات والمدير المالي والإداري وبعض الأفراد الآخرين على حسب الحاجة، وقد تكون هذه اللجنة دائمة بإحدى إدارات الشؤون المالية، وتقوم اللجنة بوضع الخطوط الرئيسية للموازنة في شكل منشور عام للجهات التنفيذية لتحديد المقترنات والبدائل المختلفة لتنفيذ المقترنات.¹

2. دراسة السوق²:

تعلق دراسة السوق بتحديد ما سيبيّعه المشروع من السلع ولمن سيّباع وأين سيّباع وكيف سيّباع وسعر البيع.

وتتضمن دراسة السوق النواحي التالية:

أ- **دراسة السلع:** دراسة تفصيلية وبكافحة الأنواع التي يقوم المشروع بإنتاجها وتوزيعها والتعرف على أرقام المبيعات السابقة والحالية والربح المتوقع من كل منها، والتمييز بين السلع ذات العائد الكبير والسلع ذات العائد القليل ومعدل دوران كل سلعة من السلع وطرق التوزيع وسياساتها والتفرقة بين السلع التي تحتاج إلى جهود ترويجية والتي لا تتطلب ذلك.

ب- **دراسة المستهلكين:** تتضمن دراسة العملاء ما يلي:

- التعرف على العملاء الحاليين والمرتقبين.
- التعرف على طرق التوزيع الملائمة لفئات العملاء وكل سلعة.
- معرفة التسهيلات والخدمات المقدمة إليهم وتکاليفها.
- تحديد العوامل الداخلية والخارجية المؤثرة على سلوك المستهلك.

تفيدنا دراسة العملاء في عملية تحديد الأرباح المتوقعة من فئات العملاء حيث يتم بعدها التركيز على الفئات التي تحقق فعلاً أرباح ملموسة واستبعد بعض العملاء نهائياً أو بصفة مؤقتة نظراً لقلة الأرباح التي تتحقق من خلال التعامل معهم وقد يعود ذلك إلى ارتفاع تكاليف التوزيع أو غير ذلك.

ج- **دراسة مناطق البيع:** تهدف دراسة المناطق البيعية إلى:

¹ عبد الفتاح زكريا فريد..، 2004 - اعداد الميزانية التخطيطية، جامعة عين شمس، مصر، ص 10-11.

² العبيد علي جاسم، ظاهر عبد القدير..، 2009- أهمية وتأثير دراسة السوق في رفع كفاءة قرارات الإنتاج في الوحدة الاقتصادية "دراسة ميدانية في الشركة العامة للصناعات الكهربائية /بغداد /" للمدة 1995-2006، مجلة الإدارة والاقتصاد، العدد الثامن والسبعون، ص 135-137.

- تحديد كميات المبيعات المتوقعة من السلع والخدمات المختلفة في المناطق البيعية وتحديد الأرباح المتوقعة من كل منطقة بيعية.
- التكاليف التي يتحملها المشروع نتيجة توزيع السلع على المناطق المختلفة.
- معرفة وتحديد الجهود الإعلانية والترويجية اللازمة لكل منطقة بيعية.
- التعرف على المنافسين في كل منطقة وأثرهم على سياسات المشروع.
- تصنيف المناطق البيعية حسب أهميتها.
- تنظيم عمل مندوبيين البيع.

د- دراسة مندوبيين البيع: تهدف هذه الدراسة إلى التعرف على:

- مجهودات مندوبيين البيع وتقديرها.
- الصعوبات والمشاكل التي تواجه مندوب البيع والعمل على تحليلها.
- وضع برامج تدريبية لمندobi البيع لتحسين أدائهم واتباع الطرق المثلى للقيام بواجباتهم على أكمل وجه.

هـ- دراسة سياسات التوزيع.

و- دراسة شروط البيع والتسليم.

ز- دراسة الائتمان للعملاء.

كـ- دراسة سياسات المنافسين.

3. إعداد الميزانية: بعد دراسة السوق وتحديد الهدف من إعداد الموازنة يتم إعداد الموازنة التقديرية وعرضها على الإدارة العليا من أجل اعتمادها.

5.2.1 مدة الموازنة التقديرية للمبيعات : (Estimated Sales budget Period)

يتم تحديد مدة الموازنة التقديرية للمبيعات بناءً على نتيجة الدراسات والتحاليل التي يقوم بها الجهة المختصة في المشروع وعلى الظروف الخاصة به.

هناك عدة عوامل تؤثر في تحديد مدة الموازنة:¹

1- موسمية المبيعات (المرطبات، المكيفات).

¹ عبيدات محمد، الضمور هاني، حداد شفيق - 2013، إدارة المبيعات والبيع الشخصي، مرجع سبق ذكره، ص215-216.

2- موسمية الإنتاج وطبيعته (الزيتون).

3- طبيعة عملية شراء الخامات ومستلزمات الإنتاج.

إذا كانت المبيعات ذات طابع موسمي فلابد من تحديد مدة الموازنة على أساس طول فترة الموسم أما

إذا كانت المبيعات دائمة أو مستمرة فيتم تحديد مدة الموازنة على أساس سنوي.

بالنسبة إلى موسمية الإنتاج فلها تأثير مباشر على تحديد مدة الموازنة التقديرية للمبيعات حيث أن

موازنة الإنتاج تعتمد بشكل كبير في تمويلها على موازنة المبيعات.

وتؤثر عملية التقلب والاستقرار في شراء الخدمات الأولية الضرورية للإنتاج في تحديد مدة الميزانية حيث

تعتمد في ذلك على العادات السائدة في السوق إضافة إلى الأعراف والمنافسة والظروف الاقتصادية

العامة.

مما سبق يتبيّن لنا أن مدة الموازنة التقديرية للمبيعات تتأثّر بظروف الإنتاج والشراء والتمويل والظروف

الأخرى المحيطة بالنشاط البيعي.

الفصل الثاني



Data Mining

إن استخدام تكنولوجيا التنقيب في البيانات يوفر للمؤسسات في جميع المجالات القدرة على استكشاف والتركيز على أهم المعلومات في قواعد البيانات، كما تركز تكنولوجيا التنقيب في البيانات كذلك على بناء التنبؤات المستقبلية واكتشاف السلوك والاتجاهات مما يسمح باتخاذ القرارات الصحيحة وفي الوقت المناسب.

وسنتناول في هذا الفصل النقاط التالية:

المبحث الأول: عناصر التنقيب في البيانات وتطبيقاتها.

المبحث الثاني: تحليل سلة التسوق وقواعد الارتباط.

المبحث الثالث: أشجار القرارات.

المبحث الأول

عناصر التنقيب في البيانات وتطبيقاتها

Elements of data mining and applications

1.1.2 مقدمة : (introduction)

ادى التطور في توليد وجمع البيانات الى وجود مجموعات بيانات ذات احجام هائلة في مجالات التجارة وفي كافة فروع المعرفة العلمية، حيث وجدت المؤسسات نفسها غير قادرة على ترجمة وفهم الكم الهائل من البيانات الموجودة في مخزن البيانات (Data Warehouse) ولم تعد وسائل التحليل التقليدية الإحصائية قادرة على التعامل معها، لذا ظهرت العديد من الدراسات منذ أواخر الثمانينيات في محاولة لحل تلك المشكلات مع البحث عن حلول تجمع بين عدة تخصصات سواء كانت قواعد البيانات والذكاء الاصطناعي والحوسبة التمازجية¹، وقد تم التوصل الى التنقيب في البيانات Data mining واكتشاف المعرفة اللذين أثبتتا وجوداً كأحد الحلول الناجحة لتحليل كميات ضخمة من البيانات وذلك بتحويلها من بيانات متراكمة وغير مفهومة إلى معلومات قيمة يمكن استغلالها والاستفادة منها بعد ذلك لتصبح معرفة.

لذا يعد التنقيب في البيانات من أسرع المجالات نمواً في تخصصات علم الحاسوب الآلي، ولقد جاءت شهرته وانتشاره من الحاجة المتزايدة لأدوات تساعد في تحليل الكميات الهائلة من البيانات وفهمها، وتنتج هذه البيانات يومياً بواسطة المؤسسات المختلفة مثل البنوك، شركات التأمين، ومستودعات البيع وعلى شبكة الانترنت.

2.1.2 مفهوم التنقيب في البيانات : (Data mining concept)

يعتبر البعض التنقيب في البيانات مصطلحاً شائعاً لاكتشاف المعرفة، في حين يضع البعض التنقيب في البيانات كخطوة أساسية في عملية اكتشاف المعرفة. فقد ظهر التنقيب في البيانات في أواخر الثمانينيات والذي كان مجالاً حديثاً مع قيمة بحثية مرتفعة في دراسة قواعد البيانات والذكاء الاصطناعي وتكنولوجيا قواعد البيانات وتعلم الآلة والاحصائيات وعرض البيانات وغيرها من المجالات النظرية والتكنولوجية².

وقد عرف الباحثون تقنيات التنقيب في البيانات بعدة طرق:

¹ Zhao,Y., Chen,Y. and Yao,Y., (2006) - **User-Centred Interactive Data Mining**, Proceedings of the Sixth IEEE International Conference on Cognitive Informatics (ICCI'06):pp. 457-466.

² Han J., and Kamber M., 2012- **Data Mining: Concepts And Techniques**, Morgan Kaufmann Publishers, pp xxiii-xxiv.

- التقيب في البيانات بأنه عملية اكتشاف الارتباطات والأنمط والاتجاهات الجديدة المفيدة من خلال التدقيق في كميات البيانات الضخمة باستخدام تقنيات تمييز النماذج بالإضافة إلى التقنيات الرياضية والاحصائية¹.
- التقيب في البيانات هو استخدام تقنيات متنوعة لتحديد المعلومات أو المعرفة في قاعدة البيانات واستخراجها والاستفادة منها في دعم القرار والتنبؤ والتقدير².
- هو الانتزاع الصعب للشيء المخفى أو المجهول مسبقاً، وكذلك المعلومات المفيدة فعلاً من البيانات³.
- هو تطبيق أساليب علمية على البيانات في حل المشاكل باستخدام خوارزميات معقدة للحصول على معلومات عالية المستوى⁴.

مما سبق يمكن تعريف التقيب في البيانات كما يلي:

بأنه عملية معالجة وتحليل لمجموعة من البيانات غالباً تكون كبيرة، باستخدام وسائل آلية لاكتشاف أنماط وسلوكيات لهذه البيانات على شكل نماذج وقواعد تقييد في حل المشاكل ودعم عملية اتخاذ القرارات.

3.1.2 صعوبات التقيب في البيانات⁵ : (Data mining difficulties)

هناك الكثير من التحديات التي تواجه ظهور التقيب في البيانات ذكر منها:

- **قابلية التوسيع (Scalability)** : بسبب التقدم الذي طرأ على توليد وجمع البيانات أصبح من الشائع التعامل مع مجموعة بيانات بأحجام تقاس بالجيجا بايت أو التيرا بايت وحتى البيتا بايت فإذا كانت خوارزميات التقيب في البيانات ستتعامل معمجموعات البيانات الكبيرة، فمن الضروري ان تكون قابلة للتوسيع، حيث تستخدم خوارزميات التقيب في البيانات استراتيجيات بحث خاصة للتعامل مع مشكلات بحث درجة تعقيدها تأخذ الشكل الأسني.
- **الابعاد العالية (High Dimensionality)** : من الشائع حالياً أن نصادفمجموعات بيانات لها مئات أو الاف السمات بدلاً من بضعة سمات حيث يزداد التعقيد الحسابي مع ازدياد الأبعاد.
- **بيانات غير المتجانسة والمعقدة (Heterogeneous and complex data)** : تتعامل طرق تحليل البيانات التقليدية معمجموعات بيانات تحوي سمات من نفس النوع سواء كانت

¹ DANIEL T.L., 2005 - **Discovering Knowledge In Data An Introduction To Data Mining**, ieee computer society, pp. 213-228.

² Hancock M., 2012 – **Practical Data Mining**, taylor & Francis Group, Abingdon,UK Published by John Wiley & Sons Inc, pp. 2-4.

³ K pujari Arun., 2001- **Data Mining Techniques**, published by Universities Press (india),pp 46.

⁴ Frawley W., Piatetsky-Shapiro G., and Matheus C., 1992 - **Knowledge Discovery In Databases: An Overview**, AI Magazine, pp. 213-228.

⁵ Hancock M., 2012 – **Practical Data Mining**, Op Cit, pp 2.4

⁵ الطويل هالة، 2009 – **التقيب عن البيانات**، دار شعاع، سوريا.

مستمرة أو فئوية ومع نمو دور التنقيب في البيانات في مجال الأعمال والعلوم والدواء وغيرها، ازدادت الحاجة إلى تقنيات يمكنها التعامل مع سمات غير متجانسة.

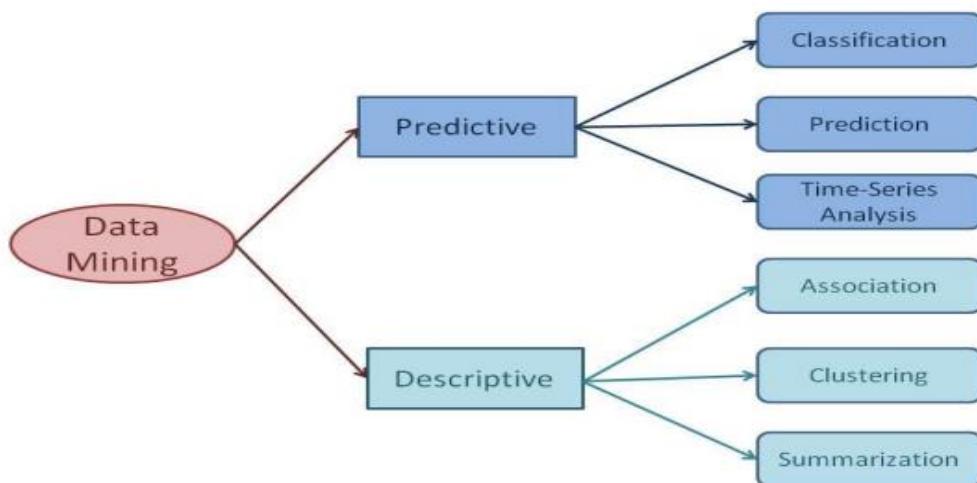
4.1.2 مهام التنقيب في البيانات : (Data mining tasks)

نقسم المهام الأساسية للتنقيب في البيانات كما يلي¹:

1. **التحليل الوصفي أو الاستكشافي (Exploratory Data Analysis)**: يقوم على وصف واستكشاف وفهم الخصائص العامة لمجموعة البيانات للوصول لنماذج مفيدة.

2. **التحليل التنبؤي (Predictive Analysis)**: يقوم بالاستفادة من البيانات الحالية كمشاهدات لتكوين نماذج قادرة على التنبؤ.

ويمكن أن يكون هناك نوع ثالث من التحليل وهو التحليل المدمج Merged Analysis والذي يعبر عن مجموعة مدمجة من التحليل الوصفي والتحليل التنبؤي، وبينما يركز الهدف التنبؤي على الدقة في المقدرة على التنبؤ، يركز الهدف الوصفي على فهم البيانات التي تولد المعالجة التنبؤية².



الشكل رقم 2 مهام التنقيب في البيانات

المصدر: Liao T., Triantaphyllou E, 2007 - Recent Advances In Data Mining Of Enterprise Data Algorithms And Applications, series on computers and Operations Research, vol 6, world scientific, London.

¹ Liao T., Triantaphyllou E., 2008 - **Recent Advances In Data Mining Of Enterprise Data Algorithms And Applications**, series on computers and Operations Research, vol 6, world scientific, London, pp 6-9.

² Larose D., Larose C., 2005 - **Discovering Knowledge In Data An Introduction To Data Mining**, Op Cip, pp277.

5.1.2 منهجيات التنقيب في البيانات : (Data mining method)

يمكن للعديد من المشاكل المتعلقة بعملية اتخاذ القرارات في أغلب المجالات أن يتم نمذجتها وفق عدة

أساليب رئيسية وهي :

- **الوصف Description:** تركز هذه المنهجية على وصف وتشخيص هيكلية وسلوك البيانات فيما بينها بأسلوب مفهوم واضح، وكمثال على ذلك الفجوة بين الجنسين في السياسة الأمريكية حيث شاهد كيف النساء يدعمن الديمقراطيين بأعداد أكبر من الرجال حيث يمكن هذا أن يثير قراراً كبيراً من الاهتمام دراسة إضافية من جانب الصحفيين وعلماء الاجتماع والاقتصاد وعلماء السياسة ناهيك عن المرشحين للمناصب العامة.¹
- **التنبؤ Prediction:** تركز هذه المنهجية على التنبؤ بقيمة دالة الهدف في المستقبل بناء على متغيرات تنبؤية² مثل توقع الزبائن الذين من الممكن أن يغادروا خلال ستة أشهر أو التنبؤ بحجم المبيعات خلال الأشهر القادمة.
- **التصنيف Classification:** تعتبر هذه المنهجية من أهم منهجيات التنقيب في البيانات والتي تقوم على إسناد البيانات إلى الفئات في دالة الهدف بناء على المتغيرات المستقلة وبالتالي تقوم على توصيف البيانات تبعاً لفئة دالة الهدف، كما يمكن تكيف هذه المنهجية لأداء مهمة التنبؤ بحيث تكون الفئة في دالة الهدف معروفة بشكل مسبق بالنسبة للبيانات المستقلة كبيانات تنبؤية، كما في حالة تصنيف مخاطر إعطاء القروض للأشخاص هل هو منخفض أو متوسط أو عالي الخطورة.³
- **التقدير Estimation:** تعتبر هذه المنهجية قريبة جداً من منهجية التصنيف لكنها لا تعطى في دالة الهدف كفئات بشكل مطلق بل تعامل معها كمتغيرات مستمرة ضمن مجالات معينة، وتقوم على تحديد قيمة دالة الهدف الرقمية بشكل تقريري باستخدام مجموعة من المتغيرات التنبؤية كما في حالة تقدير ضغط الدم للمريض بناء على قراءة بيانات تشخيص حالة المريض.
- **العنقادة Clustering:** تقوم هذه المنهجية على تجميع البيانات ضمن مجموعات مختلفة تمتلك خصائص مشابهة تسمى عناقيد، وتحتاج هذه المنهجية عن التصنيف بعدم وجود متغير هدف أي عدم معرفة هذه المجموعات بشكل مسبق، كما في حالة تجميع العملاء ضمن مجموعات.

¹ Berry M., Lionoff G., 2004 – Data Mining Techniques For Marketing, Sales And Customer Relationship Management, Wiley, Indiana, Canada, pp8.

² Larose D., Larose C., 2005- Discovering Knowledge In Data An Introduction To Data Mining, Op Cit, pp 8-13.

³ Berry M., Lionoff G., 2004 – Data Mining Techniques For Marketing, Sales And Customer Relationship Management, Op Cit, pp 8-10.

- قواعد الاقتران **Association rules**: تهدف هذه المنهجية إلى اكتشاف العلاقات والترابط بين عناصر البيانات مكونة بذلك علاقات وقواعد بين المتغيرات، كما في حالة تحليل سلة التسوق.¹

6.1.2 تطبيقات التنقيب في البيانات (Data mining applications)

يلعب التنقيب في البيانات دوراً أساسياً في البنك، والتأمين والنقل والتجارة، ويمكن للتنقيب في البيانات حل كثير من المشاكل الحسابية، ويساعد على زيادة الأرباح وصنع قرارات حكيمة. وتعتبر العمليات التجارية من أكثر التطبيقات أهمية في مجال التنقيب عن البيانات، لأن العمليات التجارية بها الكثير من بيانات المبيعات، مثل تسجيلات الشراء الخاصة بالمستهلكين، ومعلومات المستهلكين، ومعلومات الخدمة وغيرها. يمكن للشركات استخدام البيانات لتصنيف المستفيدين منمجموعات المستفيدين الأساسية وإيجاد الخصائص المشتركة للمستفيدين ورغباتهم المستقبلية، وتقدم منتجات كافية وخدمات تلبي رغباتهم. وعند استخدام تطبيقات التنقيب في البيانات، يجب اختيار الخوارزمية المناسبة، من أجل الحصول على المعرفة المناسبة. وبما أن التنقيب في البيانات يقوم بعدة مهام مثل: تجميع البيانات وتخزينها وتنظيمها لذا فهو يستخدم في عدة مجالات مثل الطب والمالية والذكاء الصناعي والقانون والدفاع والتعليم وعمليات التحكم وغيرها، وتستخدم معظم التطبيقات التنقيب في البيانات للدعائية والتسويق والمبيعات، كما يمكن استخدامه في التشخيص²، ومن الأمثلة على هذه التطبيقات:

- مكاتب الائتمان على القروض: تعتمد على ملاحظات الأفراد المشابهين في نماذج الشراء، والدخل، والقروض، كما يمكن تطوير وإنشاء تقارير موجزة عن الزبائن المهمين وعن بطاقات الائتمان.³
- السوبر ماركت: حيث تفيد دراسة بيانات السوبر ماركت في معرفة العلاقات بين المواد وترتيب هذه المواد بطريقة معينة لتلبية رغبات الزبائن.⁴
- شركات الأدوية: تقوم بتحليل الوصفات الطبية لإرسال الأدوية للزبائن المستهدفين، وإيجاد أفضل السلوكيات لاستعمال الخدمات والسلع.⁵
- وكالة الاستخبارات: تستعرض نماذج الإنفاق وبيانات السفريات للكشف عن السلوكيات غير الطبيعية عن موظفيها.⁶

¹ Sondwale P., 2015 - **Overview Of Predictive And Descriptive Data Mining Techniques**, International Journal of Advanced Research in Computer Science and Software Engineering, Vol 5, pp 264.

² Thuraisingham B., 2000 - **A Primer For Understanding And Applying Data Mining**, IT Pro IEEE Xplore, pp 28-31

³ Bharati M., Ramageri., 2010 - **Data Mining Techniques And Applications**, Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305.

⁴ Zhu C., Xue M., 2009 - **The Application Of Data Mining In The Decision Of Supermarket Extension And Businesses Expansion Based On Evolutionary Computation**, iee xplore, 10.1109/PACCS.2009.91.

⁵ Ranjan J., 2007 - **Applications Of Data Mining Techniques In Pharmaceutical Industry**, Journal of Theoretical and Applied Information Technology, pp 65.

⁶ Pejic Bach M., 2003 - **Data Mining Applications In Public Organizations**, conference paper, IEEE xplore.

- طبيب التحاليل: يقوم بتحليل صور الأشعة السينية لاكتشاف الأنماط غير الطبيعية.¹
- نظام حجز الطيران: يستخدم معلومات حول نماذج السفر والاتجاهات لتحقيق أقصى قدر من استخدام المقاعد.²
- البنوك: تستخدم التنقيب في البيانات للحصول على بيانات تساعد على جذب الزبائن.³

وعلى الرغم من استخدام هذه التطبيقات لبعض الوقت التحليل الاحصائي، إلا أن الموظفون مؤخرا بدأوا باستخدام تكنولوجيا التنقيب في البيانات لتحليل البيانات وإنشاء علاقات متبادلة ووضع التنبؤات.

7.1.2 خطوات التنقيب في البيانات (Data mining steps):

إن التنقيب في البيانات لا يقتصر على التحليل الآلي وتطبيق الخوارزميات بل يشمل مجموعة واسعة من العمليات على شكل خطوات متتالية، حيث تعتمد كل خطوة على الخطوة السابقة لها، كما أنه يمكن الرجوع عند كل خطوة إلى الخطوة السابقة، وبشكل عام تتتألف عملية التنقيب في البيانات من ست خطوات رئيسية هي:

- **فهم طبيعة العمل (Business Understanding):** بداية لابد من فهم المجال الذي يتم فيه تطبيق التنقيب في البيانات، وتحديد الهدف منه ووضع الخطة الأولية للوصول اليه، وتحويله الى مشكلة تم توصيفها عبر التركيز على النقاط الرئيسية ومعرفة العقبات التي قد تواجهنا.⁴
- **فهم البيانات (Data Understanding):** بعد تحديد المشكلة للوصول للهدف المطلوب، لابد من فهم البيانات المتوفرة، وتشمل هذه المرحلة على جمع ووصف البيانات والتحقق من مدى جودتها وتوفرها وكفايتها، ومعرفة العيوب التي تعاني منها، وتحديد خصائص هذه البيانات من حيث إمكانية تمثيلها للمشكلة المطروحة، فبعض البيانات ليس لها ضرورة، ويتم في هذه الخطوة أيضاً فهم العلاقة بين هذه الخصائص لتحديد الهدف الذي نريد الوصول اليه، إضافة إلى تقدير مدى إمكانية استخدامها والحصول عليها مستقبلاً، وقد نضطر إلى إعادة النظر في الخطوة السابقة للحصول على أفضل البيانات التي تحقق الهدف.⁵
- **تحضير البيانات (Data Preparation):** يتم في هذه الخطوة تجهيز البيانات الخام بتحديد نوعية البيانات المستهدفة سواء كانت رقمية أو إسمية، وإجراء عدد من العمليات الحسابية والمنطقية عليها، كما

¹ George Eapen A., 2004 - **Application Of Data Mining In Medical Applications**, master thesis, Systems Design Engineering, Ontario, Canada, pp 15.

² Pal J., 2011- **Usefulness And Applications Of Data Mining In Extracting Information From Different Perspectives**, annals of library and information studies vol.58, pp 7-16.

³ Bhambri V., 2011 - **Application Of Data Mining In Banking Sector**, International Journal of Computer Science and Technology india, Vol 2, pp 199-201.

⁴ Brown M., 2014 - **Data Mining For Dummies**, A JOHN WILEY & SONS, INC, Canada, pp 76-78.

⁵ Brown M., 2014 - **Data Mining For Dummies**, OP Cit, pp 76-78.

يتم النظر فيها من حيث تجانس هذه البيانات بتمثيلها على أساس سجلات، بحيث يمثل كل سجل حالة أو كيان من هذه البيانات يحوي على خصائص معينة، وإذا كان هناك دالة هدف لابد من تحديدها وذلك تبعاً للمنهجية المستخدمة في عملية التنقيب في البيانات.¹

- **النمذجة Modeling:** تعتبر هذه الخطوة متداخلة بشكل كبير مع الخطوة السابقة وتترافقان مع بعضهما، حيث يتم تحديد المنهجية التي سيتم بها نمذجة البيانات بما يناسب المشكلة المطروحة، واستخدام الخوارزميات والتقنيات المختلفة للوصول للنموذج، ولابد هنا من العودة للخطوة السابقة بحيث يكون تجهيز البيانات يناسب المنهجية والتقنية المستخدمة، للوصول للنموذج الأفضل.
- **التقييم Evaluation:** في هذه الخطوة يتم تقييم النموذج بدقة للتأكد من أنه يحقق الهدف بشكل صحيح وبدقة مناسبة، وذلك للتوصل إلى قرار بشأن استخدام النماذج الناتجة عن عملية التنقيب في البيانات، أو إعادة النظر في الخطوات السابقة، وتعتبر من أهم الخطوات للحكم على النماذج قبل تعميمها.
- **الendum** Deployment: ويتم في هذه الخطوة استخدام النموذج مع البيانات والحالات المستقبلية.²



الشكل رقم 3 خطوات التنقيب في البيانات

المصدر: Olson D. Delen D. 2008 - Advanced Data Mining Techniques, Springer publishing

¹ Kantardzic M., 2011 - **Data Mining: Concepts, Models, Methods, And Algorithms, 2nd Edition**, iee press, new jerssy, Canada, pp 26-51.

² Padhy N., 2012 - **The Survey Of Data Mining Applications And Feature Scope**, International Journal of Computer Science, Engineering and Information Technolog, Vol.2, No.3, pp 46.

المبحث الثاني

تحليل قواعد الاقتران

Analysis of Association Rules

1.2.2 مقدمة : (introduction)

أدى الاستخدام الكبير لقواعد البيانات إلى وجود مجموعات بيانات ذات أحجام هائلة في مجالات التجارة وفي كافة فروع المعرفة العلمية، حيث وجدت المؤسسات نفسها غير قادرة على ترجمة وفهم الكم الهائل من البيانات الموجودة في مخزن البيانات (Data Warehouse) ونتيجة لذلك ظهرت أدوات وخوارزميات تساعد في عملية اكتشاف وتحليل البيانات وإعطاء المستخدم المعلومات المفيدة، مثلًا في الهايبر ماركت (Hypermarket) يتم بشكل يومي جمع بيانات كثيرة عن عمليات شراء الزبائن للمواد وحركة هذه المواد وتحليلها باستخدام خوارزميات عديدة منها تحليل قواعد الاقتران.

2.2.2 تعريف قواعد الاقتران¹ : (Definition of Association Rules)

وهي طريقة بحث واسعة الانتشار تستخدم لاكتشاف العلاقات بين المتغيرات في قواعد البيانات الضخمة وذلك باستخدام خوارزميات بحث خاصة حيث يمكن تمثيل العلاقات المكتشفة بشكل قواعد الاقتران ويمكن التعبير عنها بالشكل التالي:

$$X \rightarrow Y$$

حيث أن شراء مادة X يؤدي إلى شراء مادة Y.

3.2.2 أنواع قواعد الاقتران² : (Association Rule Types)

بعد تحليل البيانات يمكن اكتشاف أنواع لقواعد الاقتران منها:²

(1) قواعد قابلة للتطبيق (Actionable Rules): وهي القواعد التي تحتوي على معلومات قيمة وذات جودة كبيرة.

¹ Berry M., & Linoff G., 2004 - **Data Mining Techniques For Marketing, Sales, And Customer Relationship Management**, Op Cit, pp 287-289.

² Gancheva, V., 2013 - **Market Basket Analysis Of Beauty Products**, Master of Science in Economics and Business, Erasmus University Rotterdam, pp 25.

- (2) قواعد بدائية (Trivial Rules): معلومات عادية معروفة جيداً من قبل متذمّي القرار .
 (3) قواعد غير مفهومة (Inexplicable Rules): وهي معلومات لا يمكن شرحها أو تفسيرها.

4.2.2 تطبيقات قواعد الاقتران (Association Rule Applications)

لقواعد الاقتران تطبيقات عديدة نذكر منها:¹

- أ- تحليل سلة التسوق (Market Basket Analysis)
- ب- التشخيص الطبي (Medical Diagnosis)
- ت- تسلسل البروتين (Protein Sequences)
- ث- بيانات التعداد (Census Data)
- ج- إدارة علاقات العملاء لبطاقات الائتمان (CRM of Credit Card Business)

وبعد تطبيق تحليل سلة التسوق من التطبيقات الشائعة والمهمة لقواعد الاقتران وذلك من أجل تحليل سلوك الشراء لدى الزبائن، حيث تتكون قاعدة بيانات سلة التسوق من مجموعة المناقلات (فواتير) وكل مناقلة عبارة عن مجموعة المواد التي يقوم الزبون بشرائها أي المناقلة هي عبارة عن فاتورة واحدة لزبون واحد.
 يتم دراسة العلاقة بين المواد ليتم استخلاص قواعد اقتران منها والتي تقييد قسم مراقبة المخزون وقسم ترويج المنتجات.

5.2.2 مفاهيم قواعد الاقتران (Association Rules Concepts)

هناك مفاهيم أساسية في قواعد الاقتران وهم:²

1.5.2.2 الدعم (Support)

نقول عن $X \rightarrow Y$ بأنها مدرومة بنسبة $S\%$, إذا كانت $S\%$ من المناقلات في قاعدة البيانات تحوي $X \& Y$ ، وهي نسبة وجود X و Y في مجموعة المناقلات T على عدد المناقلات الكلية.

2.5.2.2 الثقة (Confidence)

نقول عن $X \rightarrow Y$ أنها ذات ثقة اذا كانت $c\%$ من المناقلات في قاعدة البيانات تحوي X والتي تؤدي إلى Y .

¹ Tang, Z., & MacLennan, J., 2008 - **Data Mining With Sql Server 2008**, Wiley Publishing, pp 5-10.

² Pujari, A., 2002 - **Data Mining Techniques**, Universities Press, India, pp 71-73.

■ أهمية استخدام الدعم والثقة:¹

الدعم مقياس هام لأن القاعدة التي لها دعم منخفض جداً قد تظهر بالمصادفة، لذلك قد تكون غير مهمة بالنسبة لمتخيhi القرار. يستخدم الدعم للتخلص من القواعد غير المهمة وبالتالي في تشذيب قاعدة البيانات.

من ناحية أخرى تقيس الثقة جدارة الاستنتاج التي تم التوصل اليه من خلال قاعدة الارتباط، وبالتالي تزيد ثقة متخذ القرار بالقاعدة، علماً بأنه لا يمكن القول بأن هناك نسبة للدعم والثقة جيدة أكثر من الأخرى وذلك حسب قاعدة البيانات التي يتم التطبيق عليها واختلاف النسب بين قواعد البيانات.

3.5.2.2 العتبة : (Threshold)

في قواعد الاقتران يوجد مفهوم ما يعرف بالعتبة وذلك للدعم والثقة (Support threshold,) وهي العتبة التي إذا تجاوزناها نعتبر أنه يوجد علاقة بين المنتجات وإذا لم نتجاوزها نهمل العلاقة، يتم وضع هذه العتبة بناء على دراسة قاعدة البيانات وخبرة محلل البيانات.

4.5.2.2 مجموعة المواد :Itemset

ليكن لدينا $\{i_1, i_2, i_3, \dots, i_d\}$ مجموعة المواد في قاعدة البيانات، حيث d عدد المواد ولتكن لدينا $T = \{t_1, t_2, \dots, t_n\}$ مجموعة كافة المناقلات، حيث n عدد المناقلات (الفواتير).

تحتوي كل مناقلة t مجموعة جزئية من المواد I ، يطلق على مجموعة المواد في المناقلة الواحدة Item set فإذا كان عدد المواد في المناقلة الواحدة K مادة، فإنها تدعى مجموعة مواد مؤلفة من K مادة، يتم تمثيلها بالشكل (K-item set).

فمثلاً مجموعة المواد (milk, Pepsi, cake) تحتوي ثلاثة مواد وبالتالي تمثل بالشكل 3-item set.

6.2.2 خوارزميات قواعد الاقتران : (Association rules algorithms)

يستخدم لبناء قواعد الاقتران مجموعة من الخوارزميات من أهمها خوارزمية Apriori, FP-tree

1.6.2.2 Apriori Algorithm²

تعتبر خوارزمية Apriori الأكثر شيوعاً لإيجاد جميع المجموعات المتكررة Frequent Sets وهي أول خوارزمية مهدت الطريق لاستخدام التشذيب استناداً إلى الدعم وهي تدعى بـ Level – wise algorithm .

1.1.6.2.2 خطوات عمل الخوارزمية:

(a) يتم تحديد أعلى مستوى من K-item sets (وهو عبارة عن عدد المواد الموجودة في item set أي في المناقلة).

¹ Berry, M, & Linoff, G., 2004 - Data Mining Techniques For Marketing, Sales, And Customer Relationship Management, op cit, pp 312-315.

² Pujari, A., 2002 - Data Mining Techniques, op cit. pp 78-83.

- .1-item sets (b) نبدأ بإيجاد المجموعات المتكررة في المستوى الأول
- .Minimum _ support (c) نوجد قيمة الدعم لكل بند ونحذف البندود التي دعمها أصغر من عتبة الدعم
- 2 -item sets (d) نستخدم النتائج التي ظهرت لدينا في الخطوة السابقة لتوليد مجموعات متكررة في المستوى الثاني
- .k -item sets (e) نعيد الخطوتين السابقتين إلى أن يصل المجموعات المتكررة frequent sets إلى المستوى
- .Min _ Conf (f) نقوم بتوليد قواعد الاقتران وذلك حسب المجموعات التي تكون ثقته أكبر من

2.1.6.2.2 ميزات خوارزمية :Apriori

- (a) سهلة التنفيذ.
- (b) تقلل عدد المجموعات المرشحة التي يجب أن تأخذ بعين الاعتبار عن طريق التسذيب.
- (c) يتم استبعاد المجموعات غير المكررة Infrequent Set.

3.1.6.2.2 مساوى خوارزمية :Apriori

تتطلب العديد من عمليات المسح لقاعدة البيانات وهذا ما يشكل عبء وجهد على الحاسوب وبالتالي سيحتاج الحاسوب إلى زمن قد يصل في بعض الأحيان إلى ساعات للحصول على النتائج حسب حجم قاعدة البيانات.

2.6.2.2 خوارزمية FP Tree¹

تستخدم هذه الخوارزمية طريقة مختلفة جوهرياً عن خوارزمية Apriori لاكتشاف مجموعة بنود متكررة حيث يتم تمثيل بيانات المناقلة على شكل شجرة، وبما أن بنود المناقلات تحتوي على مواد مشتركة لذلك فإن الشجرة ستمتلك مسارات متشابكة.

1.2.6.2.2 خطوات عمل الخوارزمية:

يتم بناء الخوارزمية وفق الخطوات التالية:

الخطوة الأولى: بناء الشجرة:

يتم عمل مسح لمجموعة البيانات لتحديد الدعم لكل مادة حيث يتم استبعاد البيانات غير المتكررة ومن ثم يتم ترتيب البيانات المتكررة تنازلياً من المادة الأعلى دعماً إلى الأقل دعماً.

تقوم الخوارزمية بمرور ثاني على البيانات لبناء الشجرة حيث يتم رسم عقد ومسارات لجميع المناقلات.

الخطوة الثانية: توليد مجموعة مواد متكررة (Frequent item set)

¹ Agrawal, R., & Srikant, R., 1994 - **Fast Algorithm For Mining Association Rules**, Morgan Kaufmann Publishers. San Francisco.

يتم توليد مجموعات بنود متكررة من خلال استراتيجية (Bottom-up) أي من الأسفل إلى الأعلى (من الورقة إلى الجذر).

تستمر عملية البحث إلى أن يتم استكمال كافة المسارات حيث تقدم خوارزمية FP-tree استراتيجية (Divide and Conquer) والتي تقسم المسارات إلى مسارات جزئية أصغر من أجل معرفة جميع البنود المتكررة.

تقوم الخوارزمية FP-Tree بجمع كافة المسارات المرتبطة بالعقدة التي بنهائية الشجرة، ونتأكد من أنها تحوي على تعداد أكبر من العتبة وفي حال أنها عقدة متكررة أي أكبر من العتبة نحو مسارات الشجرة البدائية إلى شجرة شرطية، وهي مشابه لبنية شجرة FP-tree فيما عدا أنها تُستخدم لإيجاد مجموعات بنود متكررة تنتهي بعقدة معينة حيث يتم الحصول عليها وفق الخطوات التالية:

- (a) يتم تحديث تعدادات الدعم للعقد حيث يتم سطح المناقلات التي لا تنتهي بالعقدة المحددة وإعادة حساب دعمها.
- (b) نقوم بحذف العقدة المحددة لأننا حصلنا على تكرارها سابقاً ونبقي على العقد الجزئية منها بسبب الحاجة إليها لاحقاً.
- (c) نقوم بإلغاء العقد التي لا تتحقق مجموع تعداد دعم أكبر أو يساوي Min_sup .
- (d) نطبق الاختبار السابق على كافة العقد الجزئية للعقدة المحددة ونحدد التكرارات للعقد.
- (e) نعيد تطبيق الخطوات السابقة كلها على جميع العقد فينتج لدينا العقد والمسارات المتكررة.

2.2.6.2.2:FP-tree مميزات خوارزمية

- 1) يتم مسح القاعدة مرتين فقط.
- 2) تعتمد على مبدأ ضغط البيانات.
- 3) أسرع من خوارزمية Apriori
- 4) سهولة تحديث الشجرة.

3.2.6.2.2:FP-tree مساوى خوارزمية

- 1) يمكن أن تكون غير مناسبة للعمل في الذاكرة.
- 2) تتطلب وقت طويل في بنائها ولكن يتم ذلك مرة واحدة فقط.

المبحث الثالث

أشجار القرارات

Decision Trees

: (introduction) 1.3.2

تعتبر أشجار القرار من أكثر طرق التصنيف استخداماً وذلك بفضل قدرتها على إنشاء نماذج قادرة على شرح وتحليل عملية دعم القرارات المعقدة على شكل مجموعة من القرارات البسيطة قابلة للتفصير بشكل سهل¹، بالإضافة إلى أنها من الطرق التي يسهل تفديها واستخدامها لأنها لا تحتاج إلى تدريبات معقدة للحصول على النتائج لاعتمادها في عملية تقسيم الشجرة على أفضل المتغيرات².

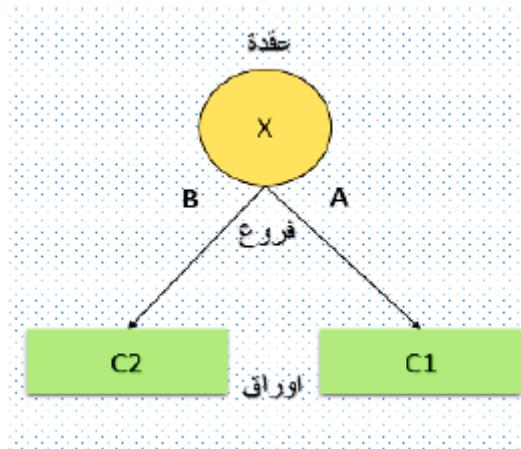
:³ (Concept of decision trees) 2.3.2

أشجار القرار هي إحدى أساليب التصنيف التي تقوم على بناء هيكل شجري لتمثيل القواعد المستخرجة من عملية التصنيف، وتتكون من مجموعة من العقد تمثل نقط اختبار، وتنتج عنها فروع تمثل مسار ناتج الاختبار، تؤدي في نهايتها إلى أوراق تمثل القرار، ويوضح الشكل التالي مكونات شجرة القرار:

¹ Alka G., Ravindra P., 2009 - **Building Privacy-Preserving C4.5 Decision Tree Classifier On Multiparties**, International Journal on Computer Science and Engineering Vol.1(3), pp 199.

² Ikonomovska E., 2012 - **Algorithms For Learning Regression Trees And Ensembles On Evolving Data Streams**, Doctoral Dissertation, Ljubljana, Slovenia, pp 23.

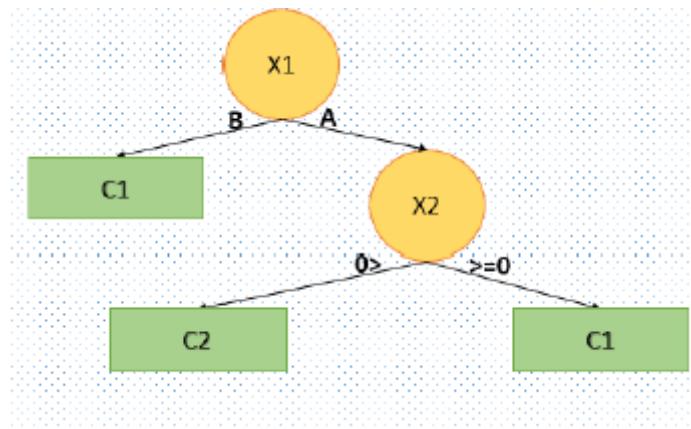
³ Mayatt G., Johnson W., 2014 - **Making Sense Of Data I. A Practical Guide To Exploratory Data Analysis And Data Mining**. 2nd Edition, Wiley, USA, pp 125.



الشكل رقم 4: مكونات شجرة القرار.

المصدر: إعداد الباحث.

وقد لا ترتبط عقد الاختبار مباشرةً بالأوراق عبر الفروع الناتجة بعد اختبار أخرى وفي النهاية تنتهي بأوراق، ويعبر عن العقد العليا عقدة أب للعقد الأدنى منها، أما العقد الدنيا فهي عقد أبناء للعقدة الأعلى منها، ويتم بناء شجرة القرار من جدول البيانات الذي يحوي عدة متغيرات مستقلة X_i ومتغير واحد هو دالة الهدف Y ، وبذلك قد يكون لكل متغير مستقل عقدة اختبار، كما هو موضح بالشكل:



الشكل رقم 5: ارتباط عقد الشجرة.

المصدر: إعداد الباحث.

وهذا يعني أن عقد الاختبار في شجرة القرار تمثل الحقول الموجودة في قاعدة البيانات أي المتغيرات المستقلة، أما الفروع التي تخرج من هذه العقد فهي عبارة عن القيمة التي تحويها هذه الحقول، وتكون هذه القيمة هي نقطة اختبار العقد بحيث تقسم مجموعة البيانات تبعاً لإحدى الفئات في دالة الهدف إلى عدة مجموعات، وفي نهاية الشجرة عند حواف الفروع توجد الأوراق أي الفئات في دالة الهدف، وتقوم الشجرة بالتصنيف عن طريق الترتيب من الأعلى إلى الأسفل وتبدأ من عقدة وحيدة تسمى عقدة الجذر تشكل مسارات إلى عقد أخرى عبر تقسيم

البيانات إلى أجزاء بشكل تكراري حتى الوصول إلى الأوراق، مكونة بذلك عدة مسارات ويفسر كل مسار على أنه قاعدة للوصول إلى فئة معينة في دالة الهدف، وبالتالي الوصول إلى القرار المطلوب وهذا يعني شجرة القرار تتشكل مما يلي:

عقدة الجذر: هي عقدة وحيدة تمثل عقدة أب لكل العقد الموجودة في شجرة القرار كاملة، وليس لها أي فروع داخلة وتخرج منها فروع ترتبط إما بعقد دخل لتشكل عقد أبناء لعقدة الجذر أو ترتبط مباشرة بورقة، وتعبر عن أهم المتغيرات المستقلة التي تقسم مجموعة البيانات بحسب الفئات.

عقد دخل: هي مجموعة من العقد الأخرى المتشكلة في الشجرة، وكل عقدة لديها فرع داخل إما من عقد الجذر أو من عقد دخل أخرى تشكل لها عقد أب وفروع خارجة ترتبط بعقد دخل أخرى تشكل عقد أبناء لهذه العقدة أو ترتبط بورقة، وعند أي عقدة من هذه العقد تتشكل شجرة فرعية مع العقد الأبناء والأوراق في المستويات الأدنى المتفرعة عنها، وتعبر هذه العقدة عن المتغيرات المستقلة في مجموعة البيانات.

الأوراق: تعبر هذه الأوراق عن نهاية المسار في الشجرة ولديها فرع داخل من عقدة دخل أو من عقدة الجذر مباشرة، ولا تخرج منها أي فروع، وتعبر عن الفئات في دالة الهدف وبالتالي القرار النهائي.

3.3.2 مميزات وعيوب أشجار القرار (Advantages and disadvantages of decision trees):

تستخدم تقنيات أشجار القرار بشكل واسع وأكبر من باقي تقنيات التصنيف الأخرى ويرجع ذلك إلى قوتها وقابليتها للفهم والتكلفة المنخفضة في عمليات حسابها وسهولة التفسير ، حيث تتمتع بالخصائص التالية:¹

- أشجار القرار قادرة على توليد قواعد مفهومة وتكون سهلة القراءة والتفسير .
- تقدم دلالة وإشارة واضحة إلى أكثر الحقول أهمية في عملية التصنيف.
- تتعامل مع المتغيرات الرقمية والاسمية.
- التعامل مع متغيرات متعددة ولو كانت بغير صلة مع دالة الهدف فأشجار القرار تقوم على اختيار أفضل انقسام لكل متغير على حد.
- السرعة في التعلم، فهي لا تحتاج إلى الكثير من العمليات الحسابية في مرحلة التدريب.²

¹ Alka G., Ravindra P., 2009- **Building Privacy-Preserving C4.5 Decision Tree Classifier On Multiparties**, International Journal On Computer Science and Engineering Vol 3, p199.

² Kuncheva L., 2014 - **Combining Pattern Classifiers Methods And Algorithms**, 2ed, Wiley, P55.

وعيوب أشجار القرار أنها أحياناً قد تنتج نماذج معقدة، بناء على البيانات الواردة في مجموعة التدريب مما يتطلب إعادة النظر في مجموعة البيانات وإعادة تشذيبها.¹

4.3.2 تطبيقات أشجار القرار : (Application of Decision Trees)

نالت أشجار القرار اهتماماً كبيراً في كثير من المجالات وأهم هذه المجالات:²

- المجالات الإدارية وخصوصاً التسويق، حيث تُستخدم في تصنيف العملاء حول استجابتهم لمنتج معين أو خدمة، بناء على بيانات سابقة للعملاء والخصائص المحددة للعملاء (العمر، الجنس، التعليم، الوظيفة، المنطقة).
- المجالات الطبية للتنبؤ بحالات مرضية كالتبؤ بحدوث نوبات قلبية وتشخيص أمراض السرطان بناء على بيانات التشخيص للمرضى، وهناك العديد من الأبحاث التي تتناول أشجار القرار في هذا المجال.
- مجالات صناعة الألعاب الحاسوبية.

إضافة إلى البحوث التي تقوم على نمذجة أشجار القرار في مجالات أخرى كالمجالات المالية.

5.3.2 استقراء شجرة القرار : (Decision Trees Induction)

غالباً ما يكون طريقة طرح سلسلة من الأسئلة من أفضل الطرق للوصول لقرار معين بحيث تؤدي الإجابة عن أحد الأسئلة إلى الانتقال لسؤال آخر وهكذا، هذه السلسلة من الأسئلة يمكن أن تتشكل ببناء شجرة القرار التي تقوم على التعلم من البيانات، حيث تقوم معظم خوارزميات بناء شجرة القرار بإنشاء هيكلية الشجرة بشكل هرمي من الأعلى إلى الأسفل والذي يبدأ من مجموعة البيانات الكلية، اعتماداً على أسلوب التقسيم والعزل وذلك بتقسـيم مجموعة البيانات الكلية عند كل عقدة إلى مجموعات فرعية وفقاً لمعايير محددة، بحيث تحدد كل مجموعة أكبر عدد ممكن من السجلات التي تحوي فئة واحدة معزولة عن الفئات الأخرى، ويتم هذا التقسيم بشكل متكرر إلى مجموعات فرعية أصغر، وفي كل مجموعة فرعية يمكنها أن تحدد فئة واحدة فقط مشكلة بذلك ورقة عند حرف الشجرة، وتسمى عملية بناء شجرة القرار باستقراء شجرة القرار.³

ويعبر استقراء شجرة القرار عن وضع القوانين العامة على أساس دراسة حالات معينة، فنموذج شجرة القرارات هو تعبير عن مجموعة من القواعد على شكل مسارات، وكل مسار فيها له دلالات معينة للوصول للقرار، ويرى

¹ Bell J., 2014- Machine Learning Hands-On For Developers And Technical Professionals, Canada, Wiley.P 46.

² Bell J., 2014 - Machine Learning Hands-On For Developers And Technical Professionals, Op,Cit,45.

³ Aggarhwari C., 2014 - Data Classification Algorithms And Applications, Taylor & Francis Group, CRC Press, pp91.

البعض أن أسلوب أشجار القرار هي الأشكال الطبيعية للتعلم الآلي، ويطلق على القواعد المستخرجة منها بقواعد الاستقرار في أدبيات التعلم الآلي.¹

وبشكل عام فإن معظم خوارزميات التصنيف الشجري تعمل وفق المراحل التالية:

(1) التقسيم **Splitting**:² وتألف عملية التقسيم مما يلي:

البداية تكون بتشكيل عقدة الجذر من مجموعة البيانات الكلية حيث تتضمن كافة السجلات، وذلك باختيار أفضل خاصية لتقسيم مجموعة البيانات إلى مجموعات جزئية بالنسبة للفئات في دالة الهدف، ويتم إنشاء عقدة الجذر كأول عقدة اختبار في هيكلية الشجرة وإضافة فروع لها بناءً على تحديد أفضل قيمة من هذه الخاصية كنقطة اختبار شرطية لتقسيم البيانات.

يتم في كل مجموعة جزئية اختيار أفضل خاصية من بينها لتقسيم مجموعة البيانات الجزئية مرة أخرى وتشكيل عقدة تعبّر عن هذه الخاصية مرتبطة بأحد فروع العقدة السابقة، وينتج عنها فروع أبناء، حيث يتم تحديد أفضل قيمة ضمنها نقسم مجموعة البيانات الجزئية إلى مجموعات جزئية أصغر، وتكرر هذه المرحلة طالما يمكن تقسيم البيانات إلى مجموعات أصغر تحتوي فئة مميزة من الفئات حتى يتم الوصول إلى الورقة.

(2) التوقف **Stopping**:³ في كل مرحلة من مراحل التقسيم يتم التحقق إذا كانت المجموعات الجزئية المتشكّلة قادرة على تحديد فئة واحدة من فئات دالة الهدف، عندها يتم الانتهاء من بناء الشجرة.

(3) التقليم **pruning**:⁴ يتم في هذه المرحلة إزالة الفروع والعقد التي تسبّب تعقيد في النموذج ولا تضيف زيادة كبيرة في قدرة التصنيف.

6.3.2 خوارزميات أشجار القرار ⁵: (Decision tree algorithms)

تطورت أشجار القرار في ميدان الذكاء الصناعي انطلاقاً من مفهوم أنظمة التعلم التي ظهرت على يد كل من (Hunt, Marin and Stone 1966) إضافة إلى الاتصال بين النهج القائم على علم الإحصاء والنهج القائم على التعلم الآلي وكانت الدوافع التي أدت إلى تطوير أشجار القرار بشكلها الحالي هي النتائج المقدمة في سلسلة

¹ Ville B., 2011 - **Decision Tree For Business Intelligence And Data Mining: Using SAS Enterprise Miner**, CRC Press, New York, USA, pp50.

² Maheshwari A., 2014 - **Business Intelligence And Data Mining**, Business Expert Press, New York, pp 72.

³ Ye N., 2013 - **Data Mining: Theories: Algorithms And Examples**, Taylor & Francis Group, pp 44.

⁴ رضا صباح، ذياب السيد، 2008 – استخدام بعض الطرق الإحصائية والتصنيف الشجري في التصنيف والتباين بافلس الشركات مالياً، مجلة العلوم الاقتصادية والإدارية، ع49، جامعة بغداد، ص303.

⁵ Ville B., 2011 - **Decision Tree For Business Intelligence And Data Mining: Using SAS Enterprise Miner**, Op Cit, pp 28- 30.

من المقالات للباحثين الاقتصاديين (Sonquist, Morgan) ضمن أبحاثهم في دراسة المؤشرات الاقتصادية والتبنّى بالأحداث الاقتصادية في معهد البحوث الاجتماعية في جامعة ميشيغان في أمريكا الشمالية، اللذان يعتبران المطوريين الأساسيين لبرامج أشجار القرار، وكان ذلك في بداية عام 1970 عندما قاموا بكتابه برنامج حاسوبي يسمى كشف التفاعل التلقائي AID. وأصبح AID أول برنامج شجرة قرار، وفي عام 1975 قام Kass بأول حل لتحسين AID وسمى النموذج الذي اقترحه CHAID نسبة لاختبار χ^2 المستخدم في استقراء شجرة القرار، وتتجدر الإشارة إلى أنه في عام 1982 قام Kass بالعمل مع إحصائي آخر (Hawkins) على وضع نهج آخر ومحسن لأشجار القرار سمي XAID.

وظهرت بعد ذلك الخوارزميات الحديثة التي تقوم على التعلم من البيانات وبعض هذه الخوارزميات تتعامل مع متغيرات مستمرة وبعضها تتعامل مع متغيرات فئوية، وكان من أهم هذه الخوارزميات خوارزمية CART وخوارزمية M5P وستنطرق في هذا البحث إلى خوارزمية M5P بالتفصيل لاستخدامها في البحث لأنها من الخوارزميات التي تعتمد على التبنّى بمتغيرات مستمرة.

وقبل الخوض في الحديث بشكل مفصل عن خوارزمية M5P يجب التنويه إلى أنواع أشجار القرار.

حيث أن هناك نوعين من أشجار القرارات وذلك حسب مخرجات الشجرة:¹

1. **أشجار التصنيف Classification Tree:** حيث أن الخاصية التي يتم التبنّى بها هي فئوية أو اسمية، وهناك الكثير من الخوارزميات التي تساعد على بناء الشجرة منها C4.5 حيث يتم تحديد أفضل الخصائص لتصنيف البيانات.

2. **أشجار الانحدار Regression Tree:** حيث الخاصية التي يتم التبنّى بها مستمرة (رقمية) وخلال فترات زمنية، وهناك نوعين من أشجار الانحدار وذلك حسب بناء الشجرة وهي إما:

- أن يتم توليد الشجرة وتكون بيانات الفروع عبارة عن قيمة رقمية مستمرة وتسمى هذه الشجرة شجرة انحدار مستمرة، حيث أن أشجار الانحدار وبخلاف اسمها لا تستخدم نماذج الانحدار الخطى في عملية التنبؤ، بل يتم التنبؤ بناءً على متوسط قيم الأمثلة التي تصل إلى العقدة.
- أن تكون الفروع في كل عقدة عبارة عن نموذج انحدار خطى وتسمى الشجرة في هذه الحالة شجرة نماذج model trees، وهي أشجار أقل حجمًا في بناء النموذج من أشجار الانحدار ولكنها أكثر قوّة منها، وهي تتبع نفس أسلوب بناء الشجرة ولكنها تختلف بأن كل عقدة تحتوي على نموذج تنبؤي

¹ Witten I.H., E Frank, Hall M.A., 2011 - **Data Mining Practical Machine Learning Tools And Techniques**, Morgan Kaufmann, Elsevier, Pp 243-245.

يتم حسابه من المتغيرات الوالصلة إلى العقدة علماً بأنه يمكن أن يكون في الشجرة عشرات النماذج وأحياناً المئات.

كما أن أشجار النماذج أكثر صعوبة على الفهم منأشجار الانحدار ولكنها تعطي نموذج أكثر دقة.

1.6.3.2 خوارزمية M5P

تعتبر خوارزمية M5 من أكثر خوارزميات تصنيف أشجار الانحدار المستخدمة، حيث تقوم ببناء الشجرة التي تتكون أوراقها من النماذج الخطية المتعددة، ويتم اختيار العقد في الشجرة بناء على السمة التي تقل الخطأ المتوقع كدالة للانحراف المعياري،¹ تم تطويرها من قبل J.R. Quinlan عام 1992²، وقد تم تطويرها من قبل Witten & Frank³ (عام 1999) وقد سميت M5P.

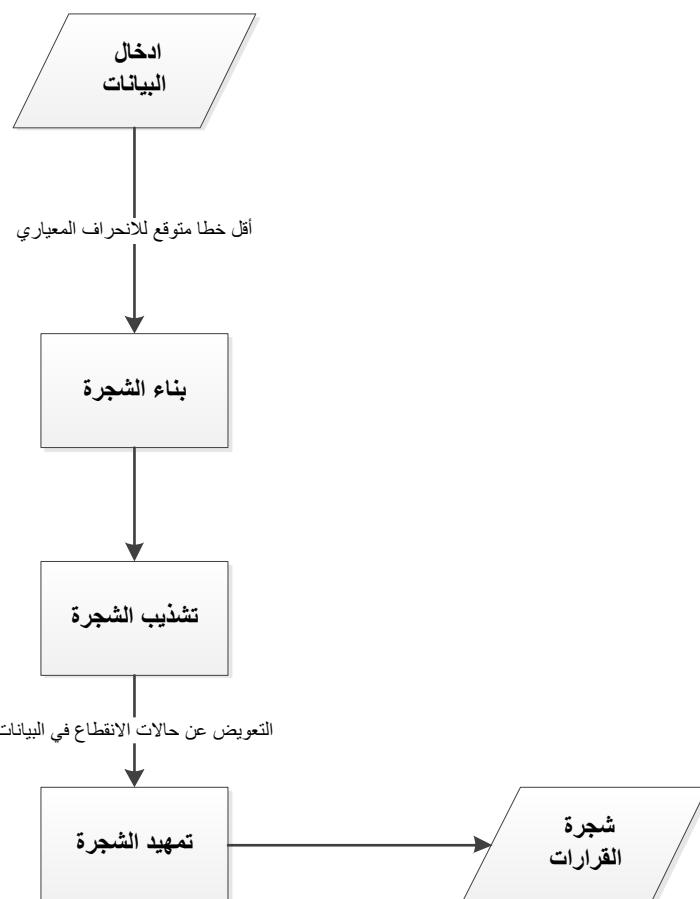
تقوم هذه الخوارزمية ببناء شجرة قرار ثنائية الفروع عن طريق تقسيم القيم في كل عقدة إلى فرعين فقط، فتقسم هذه الشجرة البيانات عند كل عقدة إلى فرعين بحيث يحتوي الطرف الأيسر القيم التي تكون أكبر من عتبة العقدة والطرف الأيمن يكون أصغر أو يساوي عتبة العقدة، وتعتمد خوارزمية M5p على مؤشر الانحراف المعياري كمعيار لعملية التقسيم وبناء الشجرة.⁴

¹ Dolado J, Rodríguez D, Riquelme J, Cuadrado J, - 2007 - A Two Stage Zone Regression Method For Global Characterization Of A Project Database, Idea Group Inc, pp 5-7.

² Chi Huynh H, 2015 - Improving M5 Model Tree By Evolutionary Algorithm, Master's Thesis in Computer Science, Halden, Norway, pp 27.

³ Singh P, Agrawal S., 2013 - Node Localization In Wireless Sensor Networks Using The M5P Tree And Smoreg Algorithms, International Conference on Computational Intelligence and Communication Networks , IEEE.

⁴ Quinlan J, 2006 – Learning With Continuous Classes. World Scientific 343-348, Singapore, Sydney, Australia.



الشكل رقم 6: آلية بناء شجرة القرارات بخوارزمية M5P.

المصدر: اعداد الباحث.

i. آلية عمل الخوارزمية:

يتم توليد شجرة الانحدار باستخدام بيانات التدريب حيث يتم حساب نموذج خطى لكل عقدة مولدة في الشجرة وذلك من خلال حساب الانحراف المعياري لقيم المستهدفة في مجموعة البيانات ويتم اختيار خاصية التقسيم بناءً على زيادة الحد من الخطأ المتوقع وبحسب المعادلة التالية:¹

$$SDR = \sigma(T) - \sum_{i=1}^k \left(\frac{N_{T_i}}{N_T} \times \sigma(T_i) \right)$$

حيث:

¹ Azadi S, Amiri H, Rakhshandehroo R., 2016 - Evaluating The Ability Of Artificial Neural Network And PCA-M5P Models In Predicting Leachate COD Load In Landfills, Elsevier Ltd, Canada.

SDR الحد من الخطأ المتوقع.

T الحالات التي تصل إلى العقدة

T_i الحالات التي تنتمي إلى الفرع M_i .

K عدد الفروع.

N_t عدد الحالات التي تصل إلى العقدة.

N_{ti} عدد الحالات التي تنتمي إلى فرع M_i .

تنتهي عملية التقسيم عندما يكون الانحراف المعياري للعقد الفرعية أقل من الانحراف المعياري للمجموعة

الأساسية أو عندما يبقى مجموعة بيانات قليلة للاختبار.¹

بعد أن يتم إنشاء نماذج الانحدار الخطي للبيانات التي تتنتمي إلى كل عقدة داخلية يتم إجراء عملية التقليم لتحقيق أقصى قدر ممكن من دقة النموذج، حيث أنه خلال هذه العملية إذا كان الخطأ المقدر لنموذج الانحدار في جذر شجرة فرعية أصغر أو مساوي لخطأ محدد مسبقاً في الشجرة الفرعية يتم استبداله بهذه الورقة، وبعد الانتهاء من عملية التقليم تأتي عملية التمهيد للقضاء على الانقطاعات في معدلات الانحدار المختلفة، حيث يتم تعديل المعادلات الخطية المتجاورة والتي تكون عبارة عن مخرجات مقدرة لمتغير إدخال واحد هي نفسها حيث يتم دمجها.

7.3.2 نقاط القوة والضعف لأشجار القرار:²

هناك نقاط متعددة تميز هذه الأشجار وهي:

1. يجمع بين قوة أشجار القرارات مع القدرة على نمذجة البيانات الرقمية.
2. تتناسب مع بعض الأنواع من البيانات أكثر من نماذج الانحدار.
3. لا تتطلب معرفة احصائية لتفسير النتائج.

أما بالنسبة لنقطات الضعف فنذكر بأنها:

¹ Onyari E, Ilunga F, 2013 - Application Of MLP Neural Network And M5P Model Tree In Predicting Streamflow: A Case Study Of Luvuvhu Catchment, South Africa, International Journal of Innovation, Management and Technology, Vol. 4, No. 1, pp 11-13.

² Azadi S, Amiri H, Rakhshandehroo R., 2016 - Evaluating The Ability Of Artificial Neural Network And PCA-M5P Models In Predicting Leachate COD Load In Landfills, Elsevier Ltd, Canada, OP.CIT.

1. ليست شائعة الاستخدام مثل نماذج الانحدار.
2. تتطلب كمية كبيرة من بيانات التدريب.
3. يمكن أن تكون أكثر صعوبة على التفسير من نماذج الانحدار.

الفصل الثالث



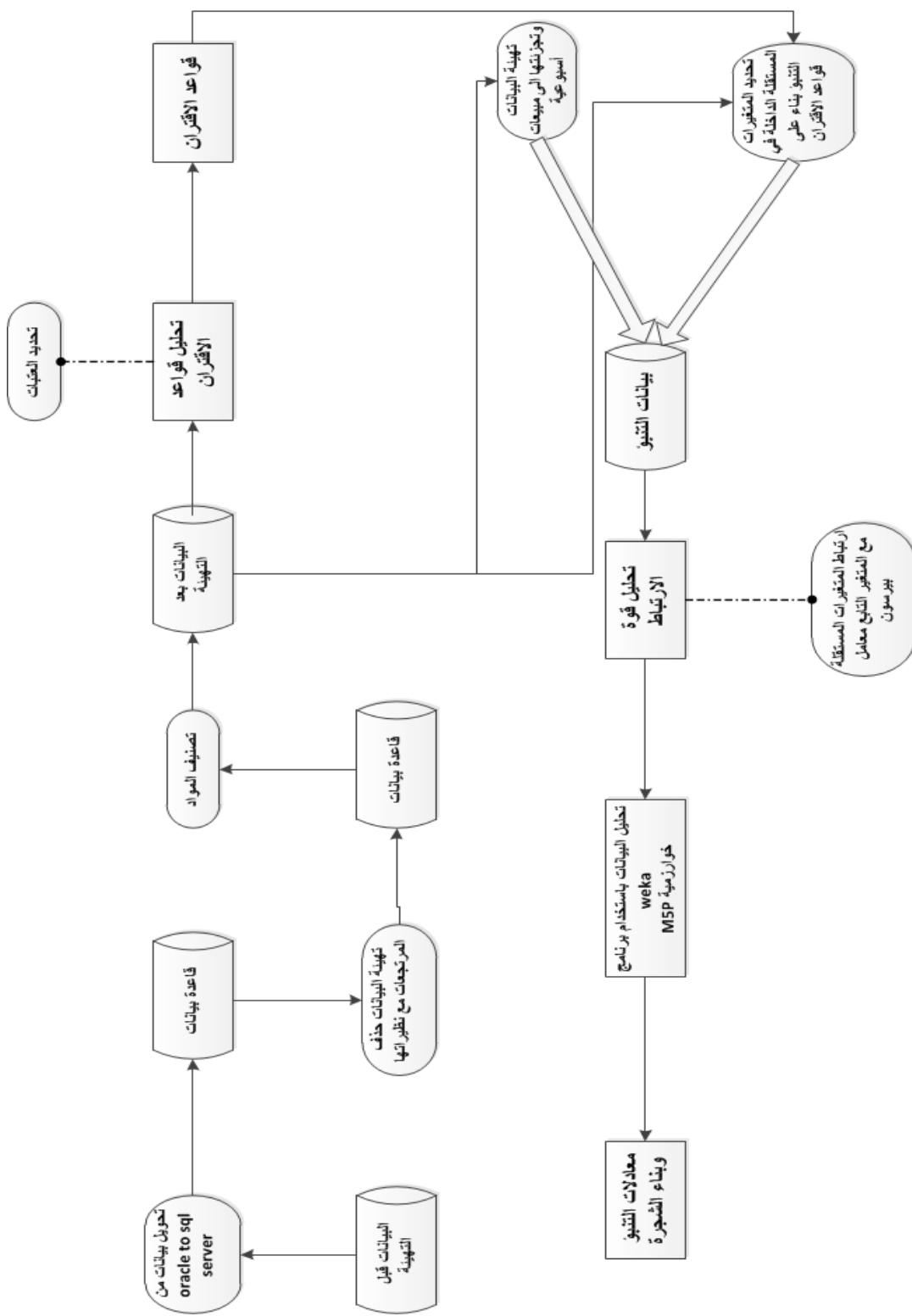
مقدمة:

في الفصل الأول تم عرض موضوع التنبؤ بالمبيعات وأهميته والعوامل المؤثرة في التنبؤ والطرق المختلفة المستخدمة في التنبؤ، أما في الفصل الثاني فتم عرض التقريب في البيانات ومنهجية التنبؤ باستخدام خوارزمية M5P والتي يمكن عن طريقها بناء نموذج للتنبؤ بالمبيعات بناء على المتغيرات المستقلة.

في هذا الفصل سيتم تطبيق الدراسة العملية من خلال التقريب في بيانات هاير ماركت في مدينة حلب، حيث سيتم تحليل بيانات سلة التسوق للهاير ماركت واستخلاص القواعد المناسبة منها ومن ثم الانتقال إلى بناء نموذج شجرة قرارات للتنبؤ بحجم المبيعات المستقبلي، مما يدعم عملية اتخاذ القرار في الوقت المناسب، وسيتم عرض ذلك من خلال المباحث التالية:

المبحث الأول: دراسة سلة التسوق.

المبحث الثاني: التنبؤ بحجم المبيعات بالاعتماد على البيانات في سلة التسوق.
وسيتم توضيح الخطوات الأساسية للتحليل من خلال المخطط التالي:



الشكل رقم 7 : مخطط الدراسة التطبيقية.

المصدر: من اعداد الباحث.

المبحث الأول

دراسة سلسلة التسويق

تم الاستعانة في هذا البحث بقاعدة بيانات لإحدى الهايبر ماركات السورية في مدينة حلب (ماركتنا) من عام 2009 وحتى عام 2010 حيث تم معالجتها بواسطة SQL server 2008.

1.1.3 مفهوم الهايبر ماركت (Haypermarket Concept)

هو عبارة عن مخزن كبير يجمع بين السوبر ماركت و المتاجر البيعية، يحتوي على مجموعة واسعة من المنتجات تحت سقف واحد (مواد غذائية، لحوم، منظفات، استهلاكية، كهربائية، ومنزلية،.....) بحيث تقوم بتلبية كافة احتياجات سوق الزبائن الروتينية في مرحلة واحدة ومكان واحد، وقد تم صياغة مصطلح الهايبر ماركت من قبل خبير التجارة الفرنسي Jacques Pictet في عام 1968.¹

2.1.3 تهيئة البيانات للتحليل:

قاعدة البيانات التي يتم التعامل بها في المول المذكور مصممة للعمل على أوراكل (Oracle 8) وقد تمأخذ نسخة من البيانات، وعند قيام الباحث باسترجاع النسخة على جهاز آخر تم مواجهة صعوبة في عملية استحضار البيانات لذلك تم الاستعانة ببرنامج مساعد لـ (Oracle 11) وهو (Toad for Oracle)، حيث تم تصدير البيانات بعد استعراضها على البرنامج المذكور على ملف بصيغة CSV، والجدول التالي يعرض جزء من البيانات المستخلصة حيث يحتوي الجدول على رقم الفاتورة ورقم المنتج والكمية المباعة في الفاتورة:

الجدول رقم 1 جزء من البيانات.

INVOICENUMBER	PRODUCTID	Total
421675	31490	2
421676	39365	2
421676	39629	2
421677	45295	1
421677	49758	1
421677	51074	1
421678	49724	1
421678	51315	1
421679	46738	3
421679	50303	1

¹ Jean Pierre., 2013 - A Forgotten Anniversary: The First European Hypermarkets Open In Brussels In 1961, Brussels Studies.

421680	46137	2
421681	168	24
421681	1204	1
421681	1287	3
421681	2132	5
421681	3297	1
421681	28634	3
421681	37438	1
421681	50380	1

مصدر: من اعداد الباحث بعد معالجة بيانات الهايبر ماركت قبل عملية التصنيف.

بعد استيراد الملف الناتج على برنامج Sql Server وتحليل البيانات تبين أن هناك كثير من المشاكل التي تم الحصول عليها في عملية التحليل وهي:

1. هناك مواد ظاهرة في قاعدة البيانات بالسابق وهي عبارة عن مرجعات بيع في فواتير، وعند القيام بعملية التحليل يعتبرها البرنامج على أنها مواد مباعة ويقوم بإدخالها في عملية التحليل ويؤدي ذلك إلى التأثير على النتائج بشكل واضح.
2. لم تقم القاعدة بإعطاء نتائج ذات معنى رغم تخفيض عتبة الدعم والثقة إلى مستويات منخفضة جداً وذلك لكثره عدد المواد في قاعدة البيانات والبالغ / 29847 / مادة.

حل هاتين المشكلتين قام الباحث بما يلي:

- إلغاء المواد المباعة بالسابق ولكن ليس بمفردها وإنما مع نظيرتها الموجبة.
- بسبب تعدد أسماء الماركات للصنف الواحد في قاعدة بيانات الهايبر ماركت مثل / سكر اليرموك، سكر الجوهرة / وهما عبارة عن مواد من نفس الصنف / سكر / أدى ذلك إلى زيادة عدد المواد في قاعدة البيانات وعند التحليل يجب علينا تخفيض عتبة الدعم إلى مستويات دنيا من أجل الحصول على نتائج لذلك قام الباحث بإعادة تصنيف لكامل مواد قاعدة البيانات وإلغاء أسماء الماركات من المواد.
- إضافة لعملية التصنيف السابقة تم تصنيف القاعدة إلى أصناف رئيسية حيث تم تقسيمها إلى: / قسم غذائية - قسم منظفات - قسم تجميل - قسم منزلية / وقد قام الباحث بدراسة قسم الغذائية فقط من القاعدة.

حيث تم معالجة جميع العمليات السابقة عن طريق خاصية Power pivot و Pivot table في برنامج

.Excel 2016

أدت عمليات المعالجة السابقة للبيانات إلى تخفيض عدد السجلات في قاعدة البيانات من / 3412333 سجل إلى / 1653839 / سجل، وإلى رفع مستوى عتبة الدعم من / 0.001 / قبل عملية التصنيف إلى / 0.05 بعد عملية التصنيف، والجدول التالي يعرض جزء من البيانات بعد تهيئتها:

الجدول رقم 2 جزء من البيانات بعد التهيئة.

اسم الصنف	الكمية	رقم المادة	رقم الفاتورة
سمن نباتي	1	1228	247085
شاي معباً	1	20755	247085
رز	2	23363	247085
حليب	3	23451	247085
بوظة	1	37567	247085
زيتون	1	37706	247085
شوربات	3	41447	247085
معكرونة	5	5717	247085
مرنديلا	1	6237	247085
مرنديلا	1	1062	247086
البان	1	28518	247086
خبز وصمن	1	28759	247086
بيض	1	28869	247086
عصائر عاديّة	24	28437	247088
البان	1	28580	247088
برغل	1	29203	247088
زبدة	1	32502	247088
سمن نباتي	1	3704	247088
فطر	4	3871	247088
رز	3	39748	247088
علكة اطفال	2	442	247088
شاي معباً	1	40074	262381
سكر	4	45983	262381

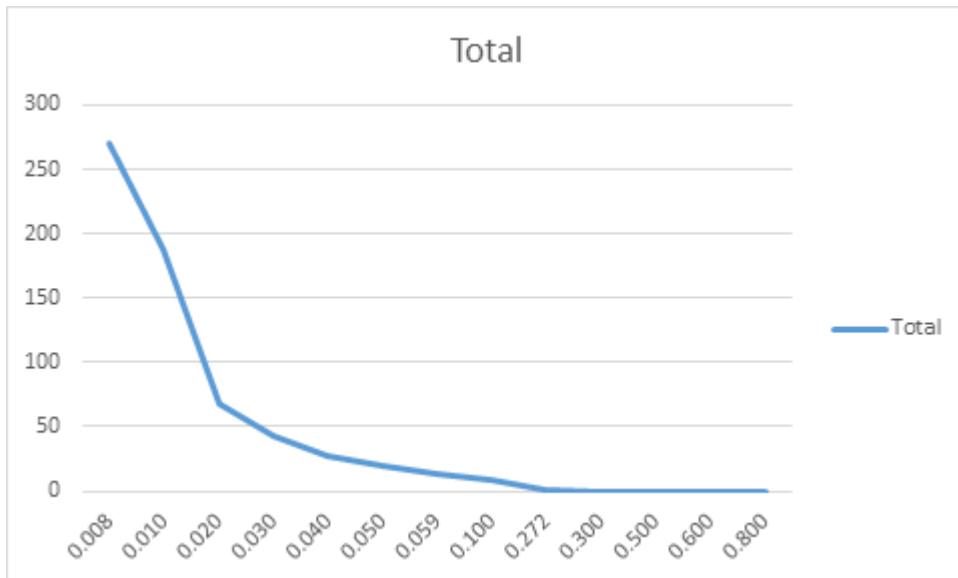
المصدر: من اعداد الباحث.

ومن الصعوبات التي يواجهها الباحث في عملية تحليل البيانات هي تحديد العتبات الدعم والثقة، حيث تختلف

هذه العتبات حسب قاعدة البيانات التي يتم تحليلها.

- في حال تمأخذ عتبات عالية يمكن أن نحصل على قواعد بدائية وشائعة ولا يمكن الاستفادة منها.
- وفي حال تمأخذ عتبات منخفضة يمكن أن نحصل على كم من القواعد الضعيفة وغير المفهومة. لذلك يجب الحذر عند تحديد هذه العتبات.

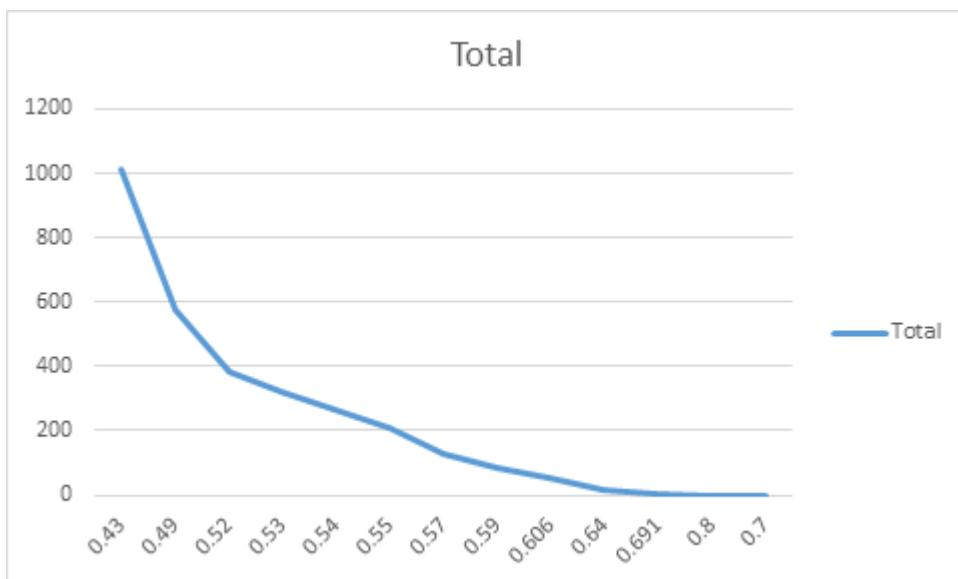
قام الباحث بتحليل قاعدة البيانات عند أكثر من مستوى للدعم والثقة، حيث تم التحليل باستخدام برنامج Sql Server Business Intelligence وتم التعبير عن النتائج التي حصلنا عليها بيانياً باستخدام Excel 2016 من خلال المخطط البياني التالي:



الشكل رقم 8: الدعم عند مستويات عتبة متعددة.

المصدر: من إعداد الباحث باستخدام برنامج الأكسل.

حيث يمثل المحور الأفقي عتبات الدعم والمحور العمودي عدد المواد الناتجة حسب عتبة الدعم، وقد تمأخذ عدد المواد الناتجة في المخطط لأن مخرجات تحليل الدعم هو عدد المواد عند كل نسبة دعم.



الشكل رقم 9: الثقة عند مستويات عتبة متعددة.

المصدر: من إعداد الباحث باستخدام برنامج الأكسل.

حيث المحور الأفقي يمثل عتبة الثقة والمحور العمودي يمثل عدد قواعد الاقتران الناتجة من التحليل، وقد تمأخذ قواعد الاقتران لأن مخرجات تحليل عامل الثقة هو عدد القواعد المتولدة عند كل نسبة ثقة.

نلاحظ من خلال الشكل رقم (8) بأنه قد تم الحصول على النتائج عند عتبة دعم 0.272 وازدادت عند تخفيف العتبة علماً بأن مستوى العتبة يختلف من قاعدة بيانات إلى أخرى حسب القاعدة محل التحليل حيث نقوم باختيار العتبة التي تقوم بإعطائنا قواعد حتى ولو كانت منخفضة.

ومن الشكل رقم (9) نلاحظ بأن نسبة الثقة لقاعدة البيانات محل الدراسة عند مستوى عتبة دعم منخفضة هي / %65 وهي تعتبر نسبة ثقة مرتفعة لقواعد الظاهره لدينا.

بعد الانتهاء من تهيئة البيانات تم البدء في بناء قواعد الاقتران عن طريق خوارزمية Apriori حيث تمت الدراسة وفق الخطوات التالية:

a. تم تحديد عتبة الدعم وعتبة الثقة للبيانات عند مستوى العتبة $\text{MIN sup} = 0.05$ و $\text{MIN conf} = 0.60$

حيث أن هذه المستوى للعتبات يعطي أفضل النتائج.

b. تم تحديد أعلى مستوى من K-item sets والبدء بإيجاد المجموعات المتكررة في المستوى الأول وحذف البنود التي دعمها أقل من عتبة الدعم.

c. تم عملية المسح لقاعدة البيانات أكثر من مرة لإيجاد جميع المجموعات المتكررة لجميع مستويات itemsets حيث يتم حذف البيانات التي دعمها أقل من العتبة.

d. توليد قواعد الاقتران من المجموعات المتكررة حسب عتبة الثقة.

وقد حصلنا على أبرز قواعد الاقتران التالية:

الجدول رقم 3 أبرز قواعد الاقتران.

الثقة	الطرف الأيسر للعلاقة		المواد	الطرف الأيمن للعلاقة
	المواد			
0.691	مايونيز	كيك	→	خبز وصمن
0.674	كورن فليكس	زبدة	→	حليب
0.670	مرتديلا		→	خبز وصمن
0.664	مرتديلا	كيك	→	خبز وصمن
0.66	بيض	كيك	→	خبز وصمن
0.654	كورن فليكس	زيوت نباتية	→	حليب

0.647	كورن فليكس	صلصات	→	حليب
0.637	كيك	البان	→	خبز وصمن
0.635	كورن فليكس	فروج	→	خبز وصمن
0.634	كورن فليكس	نودلز	→	حليب
0.633	مايونيز	شوكولا	→	خبز وصمن
0.631	كورن فليكس	كيك	→	حليب
0.625	طحين	اجبان	→	خبز وصمن
0.621	كورن فليكس	معكرونة	→	حليب
0.62	مرتديلا	حمص	→	خبز وصمن
0.618	مياه	زبدة	→	خبز وصمن

المصدر: من اعداد الباحث بالاعتماد على مخرجات SQL server

من خلال النتائج السابقة نجد بأن القواعد التي قد ظهرت لدينا ذات عامل ثقة أكبر من / 0.6 ، والتي تعني بناء على القاعدة الاولى الظاهرة لدينا أن الأشخاص الذين يشترون المايونيز والكيك يقومون بشراء الخبز معهم بنسبة 60% من المبيعات في قاعدة البيانات (أي من مجمل الزبائن الذين قاموا بشراء المايونيز والكيك).

3.1.3 فوائد نتائج قواعد الاقتران (Benefits of the association rules results) :

يمكن الاستفادة من هذه النتائج من قبل مدير التسويق أو متخذي القرار من أجل زيادة المبيعات وتلبية رغبات الزبائن والسوق ودعم قرارات التسويق من خلال:

أ- عرض البيع وحملات التسويق (البرشورات الدعائية) :

حيث يمكن مثلاً طرح بروشور دعائي بالأصناف الظاهرة في النتائج (كورن فليكس، حليب، سمن نباتي، زيت نباتي، رز) وبأسعار منافسة من الممكن أن تصل لحد التكلفة في بعض الأحيان من أجل بيع مواد أخرى على حساب هذه المواد وذلك حسب سياسة التسويق المتتبعة في الهايبر ماركت.

كما يمكن الاستفادة من هذه النتائج في تشكيل مواد العروض، أي وضع مادة ذات دعم (Support) عالي مثل (المايونيز) مع مادة ذات دعم منخفض أو مادة قليلة الحركة مثل (الخردل) مع سياسة تسعير جيدة حيث يؤدي ذلك إلى زيادة المبيعات وتقليل المواد الرائدة في الهايبر ماركت .

بـ - حزم المواد (Package) :

تراجأً الهابير ماركات في مواسم محددة إلى وضع مجموعة من المواد مع بعضها البعض كحزمة رمضان أو حزمة عيد الميلاد أو رأس السنة وبأسعار منخفضة حيث من الممكن أن تحوي هذه الحزمة على مواد ذات دعم عالي لترغيب الزبائن في شراء الحزمة ووضع بعض المواد ذات الدعم الأقل من أجل زيادة حركة هذه المواد في المخزون.

تـ - التحكم بالمخزون:

يرتبط حجم المخزون ارتباطاً وثيقاً بمستوى المبيعات ويرتفع بارتفاعها، حيث يجب أن يؤخذ بعين الاعتبار قدرة المؤسسة على تلبية رغبات الزبائن، ومن خلال نتائج تحليل سلة التسوق يمكننا تلبية هذه الرغبات وتجاوز مشكلة نقص بعض المواد في المخزون.

كما يمكن معرفة المواد التي سيحصل عليها طلب متزايد في أشهر معينة، وبالتالي القيام باتخاذ الاجراءات اللازمة لتوفير هذه المواد في الأشهر المطلوبة.

وأيضاً من التحليل نجد في حال وجود علاقة بين مادتين يتم الحفاظ على تواجد هاتين المادتين في المخزون مع بعضهما لتلبية احتياجات الزبائن.

4.1.3 نتائج المبحث الأول:

بعد الانتهاء من تفسير نتائج التحليل لا بد من ملاحظة النقاطتين التاليتين:

❖ إن النتائج التي تم الحصول عليها بمقدور خبير أن يكتشفها من دون هذا التحليل ولكن ليس بشكل دائم، فتكرار عملية التحليل بشكل دوري يمكن أن تظهر لنا نتائج أو علاقات أو ارتباطات بين المواد قد تغيب عن ذهن الخبير أحياناً.

❖ يمكن الاستفادة من هذه النتائج في حال رغبة الشركة في افتتاح أفرع جديدة.

❖ ومن الممكن أيضاً الاستفادة من تحليل قواعد الاقتران في إعداد برنامج خبير يقوم بعملية تحليل دائم للمناقلات وأمداد أصحاب القرار بمختلف المقترنات التي تساهم على دعم عملهم.

من خلال النتائج التي تم الحصول عليها من تحليل سلة التسوق سيتم اختيار المواد الظاهرة لدينا في قواعد الاقتران والاستفادة منها في عملية التنبؤ بحجم المبيعات.

المبحث الثاني

التنبؤ بحجم المبيعات بالاعتماد على البيانات في سلة التسوق

Predict sales volume based on data in your shopping cart

(Introduction 1.2.3)

بعد أن قام الباحث بتحليل بيانات سلة التسوق وحصوله على علاقات الاقتران بين مجموعة من المواد، قام الباحث باختيار عدة متغيرات منها التي تؤثر في قيمة المبيعات من أجل إدخالها في عملية التنبؤ وجميع هذه المتغيرات هي عبارة عن متغيرات مستمرة Continuous وليس فتوية، حيث استمد الباحث بيانات هذه المتغيرات من البيانات التاريخية للهايبر ماركت محل الدراسة.

من ضمن خطوات التحليل التي قام بها الباحث في هذا المبحث تحديد المتغيرات اللازمة للدراسة وتهيئة هذه البيانات لتناسب حجم المبيعات الأسبوعي ودراسة مقدار ارتباط هذه المتغيرات بحجم المبيعات الإجمالي أسبوعياً.

2.2.3 المتغيرات الداخلة في التنبؤ (Variables involved in prediction) :

قام الباحث باختيار عدد من المتغيرات المستقلة وادخالها في التحليل حيث تم اختيار أعلى المتغيرات الظاهرة لدينا وقد تم اختيار أول 8 متغيرات للمواد الناتجة عن تحليل قواعد الاقتران وأول أربع مواد ذات الدعم العالي وهي:

- **النفقات الإعلانية:** مصاريف الدعاية والإعلان المنفقة خلال هذه الفترة على عمليات ترويج المبيعات والتي تؤثر على حجم المبيعات، وتم الحصول على هذه البيانات من قواعد بيانات الهايبر ماركت الداخل في الدراسة.
- المواد ذات العلاقات مع بعضها البعض والناتجة عن عملية تحليل قواعد الاقتران، أي أكثر المواد الناتجة في قواعد الاقتران والتي تؤثر على حجم المبيعات الأسبوعي.
- المواد ذات الدعم العالي الناتجة أيضاً عن تحليل قواعد الاقتران، أي أكثر المواد ذات الدعم العالي والتي تؤثر في حجم المبيعات الأسبوعي.

3.2.3 تهيئة البيانات (Data Configuration) :

إن كامل البيانات محفوظة ضمن بيئة Sql Server وموزعة على شكل مبيعات يومية حيث قام الباحث بعدة خطوات لتهيئة البيانات عن طريق خاصية Power pivot في برنامج Excel 2016 وهي:

- تصدير البيانات من Excel 2016 إلى برنامج Sql server

- فرز (تجميع) حجم (قيمة) المبيعات الاجمالية وفق دالة الأسبوع Weeknum() على برنامج Excel 2016.
- فرز (تجميع) أحجام (قيم) المبيعات المواد الناتجة عن تحليل قواعد الاقتران وفق دالة الأسبوع Weeknum() على برنامج Excel 2016.
- تجميع أحجام (قيم) المبيعات في الخطوتين السابقتين وتهيئتها بشكل أسبوعي باستخدام خاصية Pivot Table.

4.2.3 تخفيف عدد المتغيرات باستخدام تحليل الارتباط (معامل بيرسون):

قام الباحث من أجل التأكيد من أن المواد تؤثر في حجم المبيعات بدراسة مقدار ارتباط حجم مبيعات كل مادة مع حجم المبيعات الإجمالي أسبوعياً حيث:

تمأخذ حجم مبيعات كل مادة خلال الفترة المذكورة ودراسة قوة ارتباطها مع حجم المبيعات الاجمالية وفق معامل ارتباط بيرسون عن طريق برنامج SPSS، حيث أن عدد المتغيرات قبل التخفيف 17 متغير، وفيما يلي جزء من بيانات التنبؤ موضحة بالجدول:

الجدول رقم 4 جزء من البيانات المهنية للتتبُّع بأحجام المبيعات الأسبوعية.

Week	sales total	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	Ad
1	1607382	121509	28130	9262	66254	48461	8668	30053	24113	15152	9378	5717	35331	12150	15180	7650	3858	39804
2	3594790	271746	62911	20715	148172	108380	19385	67211	53927	33887	20972	12785	79015	9090	12120	8190	6288	89018
3	3128822	236521	54756	18030	128966	94331	16872	58499	46936	29494	18254	11128	68773	7110	17370	6024	3568	77480
4	2760410	208672	48309	15907	113780	83224	14885	51611	41410	26021	16104	9817	60675	5400	27570	1620	4374	68356
5	2520342	190524	44108	14523	103885	75986	13591	47122	37808	23758	14704	8964	55398	5910	21600	4248	2837	62412
6	2828760	213838	49505	16301	116597	85285	15254	52888	42435	26665	16503	10060	62177	8460	18570	9198	3466	70049
7	2984474	225610	52230	17198	123016	89979	16094	55800	44771	28133	17412	10614	65600	2190	22740	4956	1410	73905
8	3196266	241620	55937	18418	131745	96365	17236	59760	47948	30130	18647	11367	70255	10920	19380	8544	6600	79150
9	2815100	212806	49266	16222	116034	84873	15180	52633	42230	26537	16424	10012	61877	7920	21390	4926	8594	69711
10	2836328	214411	49638	16344	116909	85513	15295	53030	42549	26737	16547	10087	62344	7770	23220	5826	4739	70236
11	2852937	215666	49928	16440	117594	86014	15384	53340	42798	26893	16644	10146	62709	10350	18900	5730	3881	70648
12	3190898	241214	55843	18387	131524	96203	17207	59659	47868	30079	18616	11348	70137	6240	19890	9384	2461	79017
13	2893005	218695	50629	16671	119246	87222	15600	54090	43399	27271	16878	10289	63589	8070	25920	2460	5040	71640
14	3054429	230898	53454	17601	125899	92088	16471	57108	45820	28793	17820	10863	67138	4590	22590	7278	4234	75637
15	3679085	278118	64386	21200	151647	110921	19839	68787	55191	34681	21464	13085	80868	5880	17370	5694	2821	91106
16	6406943	484329	112126	36920	264085	193164	34549	119789	96113	60395	37379	22786	140827	2520	24060	3804	1200	158656
17	4554650	344306	79709	26246	187736	137319	24561	85157	68326	42935	26572	16198	100113	8340	16830	4704	1666	112788
18	4489841	339407	78575	25872	185065	135365	24211	83945	67354	42324	26194	15968	98689	1800	16980	7884	2438	111183
19	4236359	320245	74139	24412	174617	127723	22844	79206	63551	39934	24715	15066	93117	6150	3012	7596	2367	104906
20	4067553	307484	71185	23439	167659	122633	21934	76050	61019	38343	23730	14466	89407	8730	13944	7374	1377	100725
21	4007216	302923	70129	23091	165172	120814	21609	74922	60114	37774	23378	14252	88080	6930	10908	8982	2396	99231
22	3928023	296937	68743	22635	161907	118427	21182	73441	58926	37028	22916	13970	86340	7170	8970	6666	2637	97270
23	4259573	322000	74545	24546	175573	128422	22970	79640	63899	40153	24851	15149	93627	10110	8760	4974	735	105480
24	3903578	295089	68315	22494	160900	117690	21050	72984	58559	36797	22774	13883	85802	5760	13416	6396	1023	96665
25	4380966	331177	76670	25245	180577	132082	23624	81910	65720	41297	25559	15581	96295	9240	18546	5310	6796	108487

المصدر: من اعداد الباحث.

تم تحليل قوة الارتباط بين المتغيرات المستقلة وحجم المبيعات الإجمالي (المتغير التابع)، وقد حصلنا على النتائج التالية موضحة بالجدول رقم (5) :

الجدول رقم 5 جزء من النتائج الظاهرة في تحليل معامل بيرسون.

		المبيعات الإجمالية	أجبان	البان	بيض	حليب	صمغ
المبيعات الإجمالية	Pearson Correlation	1	.848**	.815**	.762**	.835**	.831**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N	105	105	105	104	105	105
أجبان	Pearson Correlation	.848**	1	.926**	.934**	.987**	.969**
	Sig. (2-tailed)	.000		.000	.000	.000	.000
	N	105	105	105	104	105	105
البان	Pearson Correlation	.815**	.926**	1	.898**	.939**	.962**
	Sig. (2-tailed)	.000	.000		.000	.000	.000
	N	105	105	105	104	105	105
بيض	Pearson Correlation	.762**	.934**	.898**	1	.947**	.943**
	Sig. (2-tailed)	.000	.000	.000		.000	.000
	N	104	104	104	104	104	104
حليب	Pearson Correlation	.835**	.987**	.939**	.947**	1	.974**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	105	105	105	104	105	105
صمغ	Pearson Correlation	.831**	.969**	.962**	.943**	.974**	1**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	105	105	105	104	105	105
عصير	Pearson Correlation	.865**	.956**	.919**	.887**	.952**	.936**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	105	105	105	104	105	105
كورن	Pearson Correlation	.856**	.944**	.950**	.903**	.942**	.951**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	105	105	105	104	105	105
مرتدية	Pearson Correlation	.777**	.942**	.924**	.904**	.949**	.951**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
	N	104	104	104	103	104	104

نجد من تحليل احدى المواد (الأجبان مثلاً) بأن معامل الارتباط $R=0.848$ أي هو أقرب إلى الواحد وبالتالي قوة الارتباط قوية.

ونجد أيضاً $Sig = 0$ أي أصغر من 0.05 وبالتالي هناك علاقة معنوية بين حجم المبيعات والمادة الداخلة في التحليل.

بعد دراسة معامل الارتباط للمواد الناتجة عن تحليل قواعد الاقتران (ذات الدعم العالي و المواد التي لها علاقة مع بعضها) تم استبعاد المواد التي ليس لها قوة ارتباط مع حجم المبيعات، أما باقي المواد فتم ادخالها في معادلة التبؤ وأصبح عدد المتغيرات لدينا 13 متغير داخلاً في الدراسة.

الجدول رقم 6 جزء من البيانات بعد التهيئة للتنبؤ بأحجام المبيعات الأسبوعية.

Week	sales total	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Ad
1	1607382	121509	28130	9262	66254	48461	8668	30053	24113	15152	9378	5717	35331	39804
2	3594790	271746	62911	20715	148172	108380	19385	67211	53927	33887	20972	12785	79015	89018
3	3128822	236521	54756	18030	128966	94331	16872	58499	46936	29494	18254	11128	68773	77480
4	2760410	208672	48309	15907	113780	83224	14885	51611	41410	26021	16104	9817	60675	68356
5	2520342	190524	44108	14523	103885	75986	13591	47122	37808	23758	14704	8964	55398	62412
6	2828760	213838	49505	16301	116597	85285	15254	52888	42435	26665	16503	10060	62177	70049
7	2984474	225610	52230	17198	123016	89979	16094	55800	44771	28133	17412	10614	65600	73905
8	3196266	241620	55937	18418	131745	96365	17236	59760	47948	30130	18647	11367	70255	79150
9	2815100	212806	49266	16222	116034	84873	15180	52633	42230	26537	16424	10012	61877	69711
10	2836328	214411	49638	16344	116909	85513	15295	53030	42549	26737	16547	10087	62344	70236
11	2852937	215666	49928	16440	117594	86014	15384	53340	42798	26893	16644	10146	62709	70648
12	3190898	241214	55843	18387	131524	96203	17207	59659	47868	30079	18616	11348	70137	79017
13	2893005	218695	50629	16671	119246	87222	15600	54090	43399	27271	16878	10289	63589	71640
14	3054429	230898	53454	17601	125899	92088	16471	57108	45820	28793	17820	10863	67138	75637
15	3679085	278118	64386	21200	151647	110921	19839	68787	55191	34681	21464	13085	80868	91106
16	6406943	484329	112126	36920	264085	193164	34549	119789	96113	60395	37379	22786	140827	158656
17	4554650	344306	79709	26246	187736	137319	24561	85157	68326	42935	26572	16198	100113	112788
18	4489841	339407	78575	25872	185065	135365	24211	83945	67354	42324	26194	15968	98689	111183
19	4236359	320245	74139	24412	174617	127723	22844	79206	63551	39934	24715	15066	93117	104906
20	4067553	307484	71185	23439	167659	122633	21934	76050	61019	38343	23730	14466	89407	100725
21	4007216	302923	70129	23091	165172	120814	21609	74922	60114	37774	23378	14252	88080	99231
22	3928023	296937	68743	22635	161907	118427	21182	73441	58926	37028	22916	13970	86340	97270
23	4259573	322000	74545	24546	175573	128422	22970	79640	63899	40153	24851	15149	93627	105480
24	3903578	295089	68315	22494	160900	117690	21050	72984	58559	36797	22774	13883	85802	96665
25	4380966	331177	76670	25245	180577	132082	23624	81910	65720	41297	25559	15581	96295	108487

المصدر: من اعداد الباحث

5.2.3 تحليل البيانات عن طريق برنامج Weka : (Data analysis by Weka)

قام الباحث بتقسيم البيانات بعد التهيئة من أجل بناء النموذج إلى مجموعتين ، المجموعة الأولى هي مجموعة التدريب وت تكون من 75 % / من السجلات والمجموعة الثانية هي مجموعة اختبار وت تكون من 25 % / من السجلات، حيث تم تدريب وتعليم النموذج واختباره من خلال خوارزمية M5P لبناء شجرة قرارات باستخدام برنامج Weka 3.7.

ت تكون قاعدة البيانات من 14 متغير، 13 متغير مستقل وهي: النفقات الاعلانية المنفقة و 8 متغيرات عبارة عن حجم مبيعات مواد تم الحصول عليها من تحليل علاقات الاقتران بين المواد و 4 متغيرات عبارة عن حجم مبيعات مواد ذات دعم مرتفع والمتغير التابع وهو: حجم المبيعات الكلي.

وقد تم ترميز متغيرات المواد من A1 الى A12 وتم الترميز للنفقات الاعلانية بـ AD .

وفيما يلي جزء من البيانات التي سوف يتم تحليلها والتي هي عبارة عن 141 / سجل.

الجدول رقم 7 جزء من البيانات الأسبوعية لمبيعات التي سيتم تحليلها.

sales total	ad	A1	A2	A3	A4	A5	A6
4786402	134296	587959	138610	56850	354155	215456	42384
4066327	114092	559916.2	130978	36010	328258	233896	36598
3972937	111472	517465.3	123378	48940	327652	215670	39646
4108063	115263	542229.7	118190	48200	356361	224748	44366
4091738	114805	604273	114296	39250	333848	235122	48384
4075524	114350	636544.5	107684	40920	335209	230606	39586
4190097	117565	601982.9	103430	45920	331273	226062	45848
4094080	114871	563109.1	99196	47630	348518	232398	45952
4523393	126917	633934.6	101102	41300	336608.3	251454	56568
3984348	111792	576688.4	101982	37080	282510	231636	45786
4051007	113662	575128.9	102284	43650	304070.6	228722	47510
3364521	94401	444722.1	80123	38790	226788.3	180580	39256
3620211	101575	656384.5	131080	46840	357939.5	279792	47791
3783144	106147	261007.8	48496	22890	134278	101973	19811
3187300	89429	218290.8	42622	23015	109805	93703	15864
2760841	77463	199826.6	41559	15167	97917	76420	11811
2993325	83986	185644.3	41911	18360	100647	80326	13866
3157514	88593	203868.6	50138	18964	110015	80691	16211
3190361	89515	200637.8	50077	15421	107937	84505	13945
3020922	84761	184847.4	48554	15999	105280	76294	10299
2858923	80215	190991.8	47721	14067	107993	73379	9773

3249623	91177	185893.2	53453	16218	137970	79312	14823
3006544	84357	179236.8	46510	15387	94020	74198	11810
2717328	76242	175931	43622	12702	84146.5	69691	8842
2797099	78481	161600.7	47627	10670	92109.5	65388	9336
2837310	79609	190619.2	49498	11616	87123	67649	7803
2804733	78695	182137.1	47055	13795	87205	67480	10156
2853336	80058	181170.5	49004	14001	105067	67365	8017
2715962	76204	194115.4	45614	20692	93670	64004	9746
2565872	71993	150742.9	44363	14154	88476	60621.65	7537
2746817	25517	204790.6	48355	8871	104963	61305	7578

المصدر: من اعداد الباحث

حيث أن Sales Total: حجم المبيعات الاجمالي.

AD: النفقات الاعلانية.

A₁..... A₁₂: المواد الناتجة عن قواعد الاقتران والمواد ذات الدعم العالي.

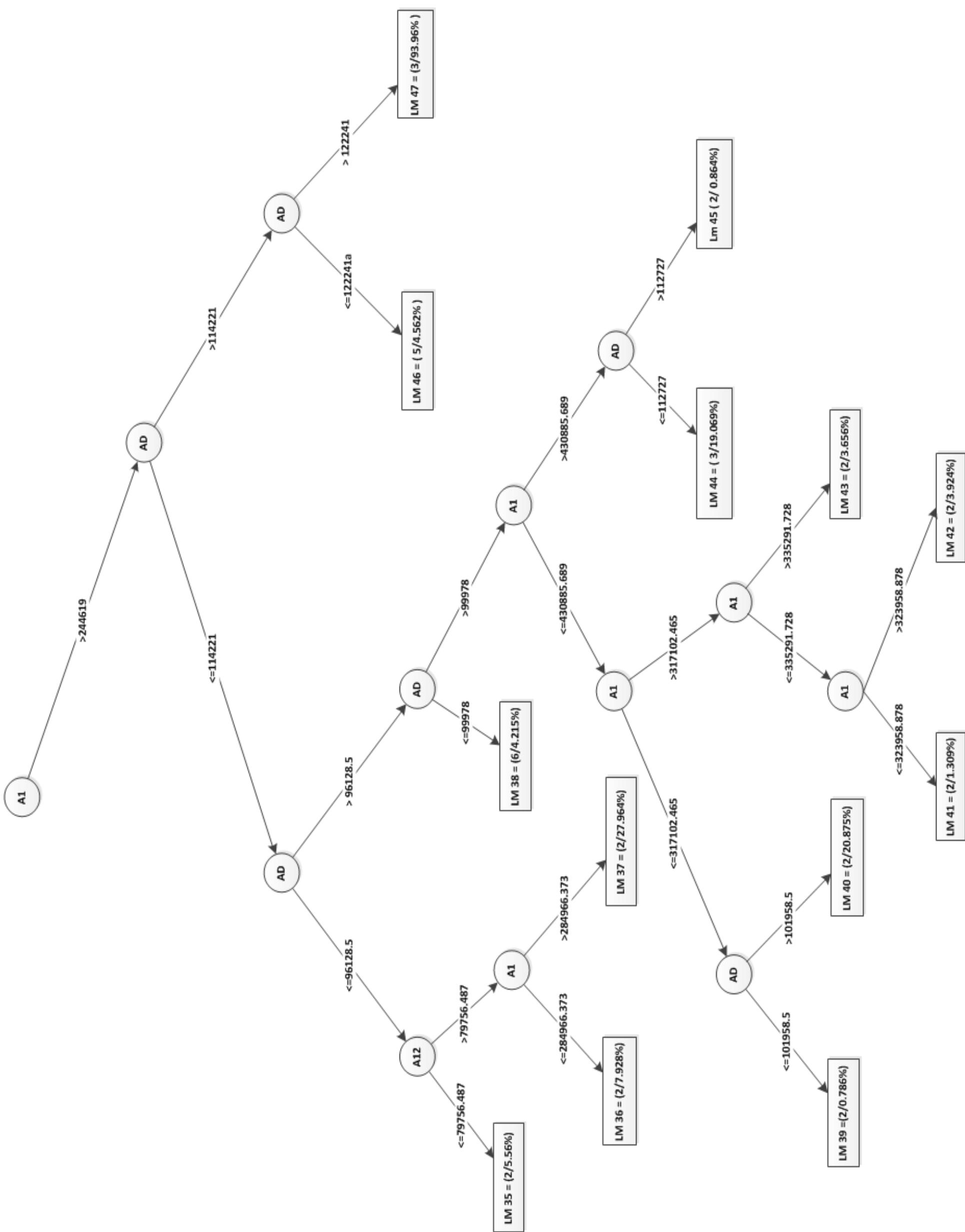
6.2.3 بناء نموذج شجرة القرارات وتقيمها (Building and Evaluating the Decision Tree Model)

تمت عملية بناء النموذج باستخدام الأداة Weka الإصدار 3.7 وهي أداة مفتوحة المصدر تعمل على منصة تشغيل Java، والتي تم تطويرها من قبل جامعة Waikato في نيوزلندا²، أما خوارزميات أشجار القرار المستخدمة فقد تم استخدام خوارزمية M5P والتي تعتبر من أهم خوارزميات التصنيف الرقمية (المتغيرات المستمرة)³. وفيما يتعلق ببناء النموذج فقد تم اتباع أسلوب التقسيم المئوي (percentage split) في عملية التدريب والاختبار ، والذي يقوم بتقسيم البيانات وفق نسبة يحددها الباحث أثناء عملية البحث حيث يتم تدريبيها ومن ثم اختبارها واطفاء النموذج النهائي، وقد قام الباحث في تقسيم البيانات إلى 75% بيانات تدريب و 25% بيانات اختبار.

وفيما يلي جزء من شجرة القرار (Model tree) التي تم الحصول عليها قبل عملية التشذيب:

² <http://www.cs.waikato.ac.nz/ml/weka/>.

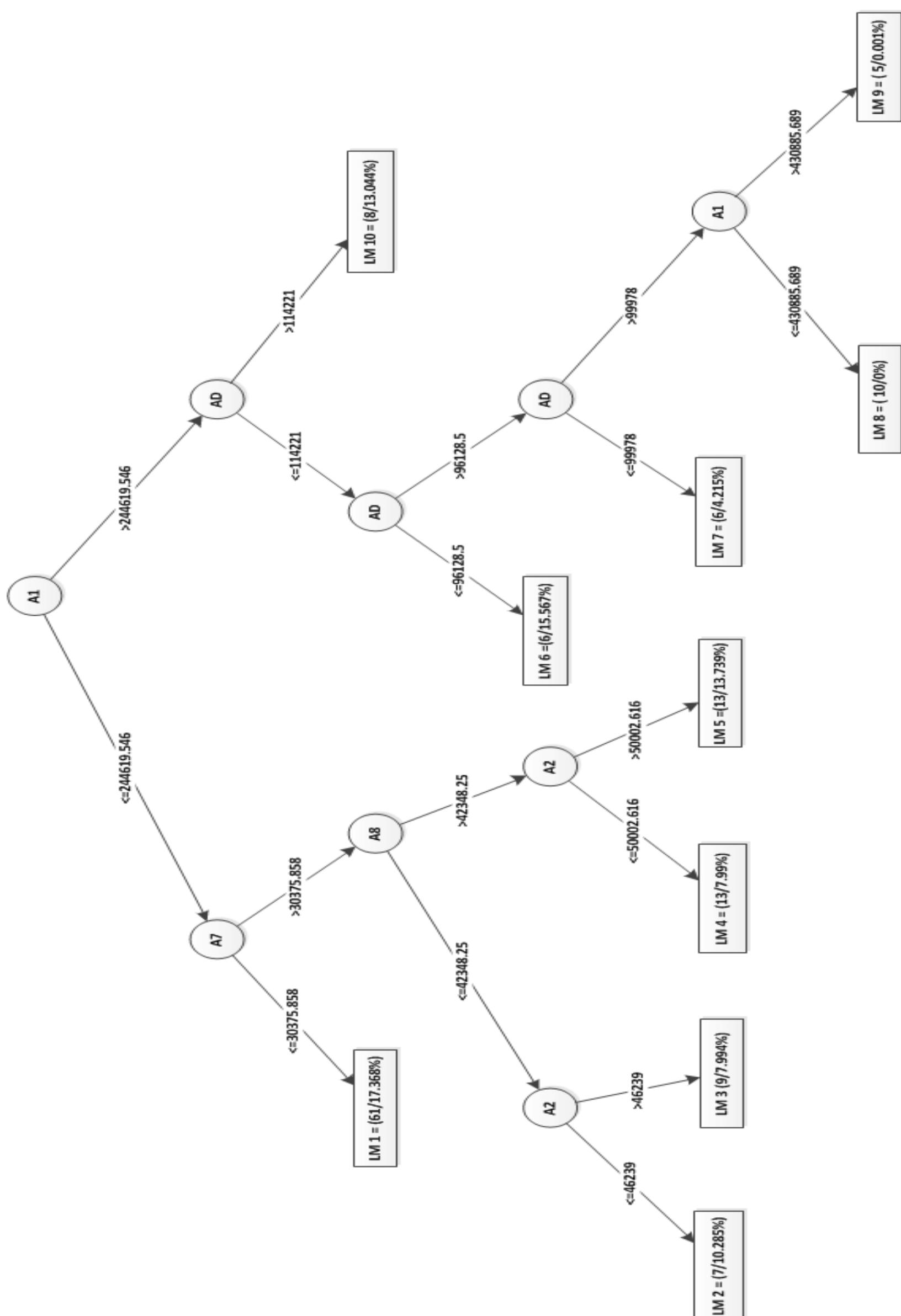
³ Deepa C, SathiyaKumari K, Pream Sudha V, 2010, Prediction of the Compressive Strength of High Performance Concrete Mix using Tree Based Modeling, International Journal of Computer Applications (0975 – 8887), pp 18-19.



الشكل رقم 10 جزء من شجرة القرارات قبل التشذيب.

. weka : مخرجات برنامج المصدر

حيث أظهرت النتائج قبل التشفيب 47 معادلة للتبيؤ، وبعد القيام بتشذيب الشجرة حصلنا على 10 معادلات للتبيؤ وفق الشجرة التالية:



الشكل رقم 11 شجرة النماذج بعد التشفيب.

المصدر: مخرجات برنامج weka.

كما حصلنا على معدلات التتبؤ التالية والخاصة بكل مسار من مسارات الشجرة العشرة:

LM num: 1 sales total = A1 * 1.4498 A2 * 27.4905 + A3 * 1.5909 - A4 * 7.2851 + A5 * 1.8666 - A7 * 0.6924 - A8 * 8.0297 + A10 * 6.5094 + ad * 5.6573 + 61807.6579 +	LM num: 2 sales total = A1 * 1.6316 A2 * 24.5328 + A3 * 2.606 + A4 * 0.4292 - A5 * 1.8666 - A7 * 0.9232 - A8 * 11.1118 + A10 * 5.444 - ad * 2.3559 + 985996.0976 +
LM num: 3 sales total = A1 * 1.6316 A2 * 24.2471 + A3 * 2.606 + A4 * 0.4292 - A5 * 1.8666 - A7 * 0.9232 - A8 * 9.0787 + A10 * 5.444 - ad * 2.3559 + 1090574.9731 +	LM num: 4 sales total = A1 * 1.6316 A2 * 28.642 + A3 * 17.6565 + A4 * 1.9779 - A5 * 1.8666 - A7 * 0.9232 - A8 * 5.9624 + A10 * 5.444 - ad * 2.3559 + 984109.0524 +
LM num: 5 sales total = A1 * 1.6316 A2 * 34.5845 + A3 * 11.7982 + A4 * 1.9779 - A5 * 1.8666 - A7 * 0.9232 - A8 * 5.9624 + A10 * 5.444 - ad * 2.3559 + 804708.1767 +	LM num: 6 sales total = A1 * 1.1243 A2 * 10.9063 + A5 * 4.4051 - A10 * 12.8478 - A12 * 1.4351 + ad * 12.9932 + 2185634.0192 +
LM num: 7 sales total =	LM num: 8 sales total =

A1 * 1.1641 A2 * 10.9063 + A5 * 4.4051 - A10 * 12.8478 - A12 * 1.2813 + ad * 17.8148 + 1749251.2107 +	A1 * 1.5508 A2 * 10.9063 + A5 * 4.4051 - A10 * 12.8478 - A12 * 1.2813 + ad * 19.2946 + 1490653.852 +
LM num: 9 sales total = A1 * 1.1028 A2 * 10.9063 + A5 * 4.4051 - A10 * 12.8478 - A12 * 1.2813 + ad * 20.2589 + 1515348.6614 +	LM num: 10 sales total = A1 * 1.6644 A2 * 10.9063 + A5 * 4.4051 - A10 * 12.8478 - ad * 26.6877 + 662426.8476 +

من خلال المعادلات نلاحظ بأنه هناك أرقام موجبة وسالبة، حيث تدل هذه الأرقام الى تصحيح الانحرافات للوصول الى رقم تتبؤى دقيق، وسنقوم بشرح المعادلات من خلال المثال التالي:

ليكن مثلاً رقم المبيعات المستهدف للمادتين $A1 = 210000$ و $A7 = 26000$ و ليكن رقم النفقات الاعلانية المقرر انفاقها خلال هذه الفترة $Ad = 19000$

نلاحظ من خلال شجرة القرار ومعادلات التنبؤ الظاهرة أننا نقوم باستخدام المعادلة الأولى والتي أعطت عامل أهمية للمتغيرين / $A1, A7$ / وذلك في حال رغبتنا باستهداف رقم مبيعات 244619 و $A1 <= 30375$ $A7 <= 244619$ لذا نقوم بالتعويض بالمعادلة التي تحقق الشرط وهي المعادلة الأولى بالشكل التالي

$$\text{LM:1 Sales total} = 1.4498 * A1 + 27.4905 * A2 - 1.5909 * A3 + 7.2851 * A4 - 1.8666 * A5 - 0.6924 * A7 + 8.0297 * A8 + 6.5094 * A10 + 5.6573 * AD + 61807.6579.$$

نعرض بالمعادلة نجد:

$$\begin{aligned} \text{Sales total} &= 1.4498 * (210000) + 27.4905 * (0) - 1.5909 * (0) + 7.2851 * (0) - 1.8666 * (0) - 0.6924 * (26000) + 8.0297 * (0) + 6.5094 * (0) + 5.6573 * (19000) + 61807.6579. \end{aligned}$$

$$\text{Sales total} = 455752.$$

7.2.3 تقييم النموذج:

تم تقييم النموذج بالاعتماد على نسبة الخطأ المطلق النسبي لنموذج شجرة القرارات المبنية بعد الاختبار باستخدام خوارزمية M5P، حيث أن برنامج Weka يقوم باعطاءنا النتيجة فقط بعد التدريب والاختبار لذلك نلاحظ أننا حصلنا على ملخص البيانات المعطى بالجدول التالي:

معايير جودة	ملخص عملية تقييم النموذج بعد الاختبار ع نسبة 75% والتدريب ع 25%	
كل ما يقرب من +1 كان جيد	0.9732	معامل الارتباط
	178600	متوسط الخطأ المطلق MAE
	263775	متوازن الجذر التربيعي للخطأ RMSE
من 0% إلى 100% كل ما يقرب من الصفر كان أفضل	%22.3352	الخطأ المطلق النسبي RAE
من 0% إلى 100% كل ما يقرب من الصفر كان أفضل	%25.7852	الجذر التربيعي للخطأ النسبي RRSE
	34	عدد سجلات الاختبار

يبين ملخص عملية التقييم أن:

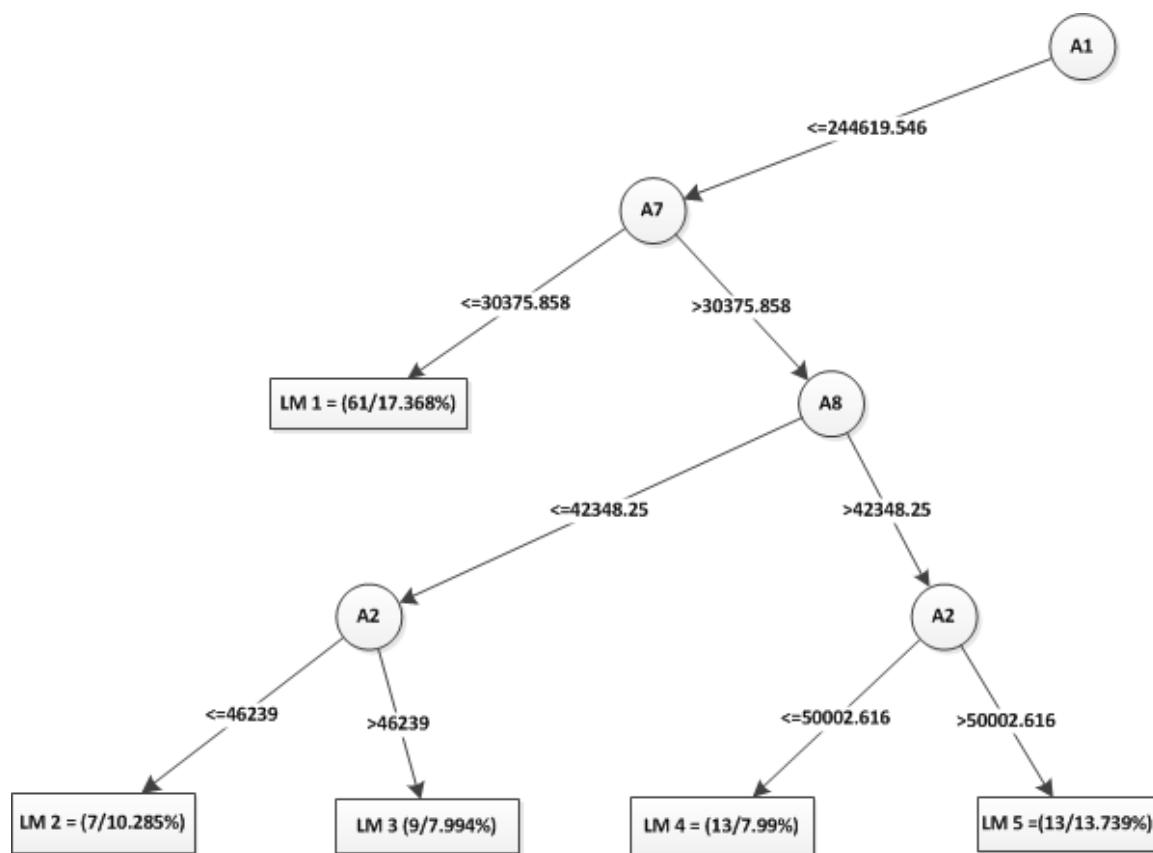
- **معامل الارتباط** هو / 0.9732 / وهو ارتباط قوي.
 - **متوسط الخطأ المطلق**: وهو الفرق بين القيم الفعلية والقيم المقدرة حيث نلاحظ بأنه / 178600 / وهو يعبر عن الفرق في حجم المبيعات ويمكن القول بأنه ليس بالرقم المرتفع مقارنة مع متوازن حجم المبيعات الأسبوعي البالغ تقريباً أربعة ملايين.
 - **الخطأ المطلق النسبي**: وهو نسبة تقسيم الخطأ المطلق على القيمة الفعلية وهو مقياس يعطينا فكرة عن مدى جودة النموذج المتبع به، نلاحظ أن نسبة الخطأ 22.3% علماً بأن هذه النسبة تكون جيدة كلما اقتربت النسبة من 0%.
- من أجل تحسين دقة النموذج لدينا قمنا بإدخال كامل البيانات في مرحلة التدريب وهم / 141 / سجل وحصلنا على النتائج التالية:

معيار جودة	% 100	ملخص عملية تقييم النموذج بعد الاختبار مع نسبة
كل ما يقرب من + 1 كان جيد	0.9807	معامل الارتباط
	132055	متوسط الخطأ المطلق
	175984	متوسط الجذر التربيعي للخطأ
من 0 % إلى 100 % كل ما يقرب من الصفر كان أفضل	% 19.0283	الخطأ المطلق النسبي
من 0 % إلى 100 % كل ما يقرب من الصفر كان أفضل	% 19.8536	الجذر التربيعي للخطأ النسبي
	141	عدد سجلات

من النتائج الظاهرة لدينا نجد بأن نسبة الخطأ المطلق النسبي قد أصبحت / 19 / % هذا يعني بأنه النسبة قد تحسنت مما يعطينا فكرة عن جودة النموذج الذي لدينا.

8.2.3 هيكلية شجرة القرارات:

بعد أن تم تقييم النموذج وتبين قدرته على التنبؤ بحجم المبيعات وبدرجة عالية من الفعالية نقوم فيما يلي بعرض هيكلية شجرة القرارات المستخرجة من عملية بناء النموذج:



الشكل رقم 12 جزء من شجرة النماذج بعد التشذيب.

المصدر: مخرجات برنامج weka

يوضح الشكل السابق عقدة الجذر (A1) والتي تمثل حجم مبيعات مادة الأجبان والتي تعتبر من أهم المتغيرات التي تقسم البيانات إلى فئات مكونة بذلك مجموعة من القواعد للحصول على معادلات للتنبؤ، فنحصل على أول معادلة للتنبؤ:

$$\text{IF } (A1) \leq 244619 \text{ And } (A7) \leq 30375 ; \text{LM1}$$

ونلاحظ أيضاً بأن المستطيل (A1) حيث يدل الرقم 61 على عدد السجلات التي تحقق الشرط السابق ويدل الرقم 17.368 % على نسبة الخطأ لهذه السجلات.

ونجد هنا أهمية المتغير (A8) والتي تمثل حجم مبيعات مادة العصائر والذي قسم مجموعة البيانات وحصلنا من خلاله على أربع معادلات للتنبؤ:

$$\text{IF } (A1) \leq 244619 \text{ And } (A7) > 30375 \text{ And } (A8) \leq 42348 \text{ And } (A2) \leq 46239 ; \text{LM2}$$

$$\text{IF } (A1) \leq 244619 \text{ And } (A7) > 30375 \text{ And } (A8) \leq 42348 \text{ And } (A2) > 46239 ; \text{LM3}$$

وبنفس الأسلوب يمكن استخراج باقي معادلات التنبؤ.

$$\text{IF } (A1) \leq 244619 \text{ And } (A7) > 30375 \text{ And } (A8) > 42348 \text{ And } (A2) \leq 50002 ; \text{LM4}$$

$$\text{IF } (A1) \leq 244619 \text{ And } (A7) > 30375 \text{ And } (A8) > 42348 \text{ And } (A2) > 50002 ; \text{LM5}$$

$$\text{IF } (A1) > 244619 \text{ And } (AD) \leq 114221 \text{ And } (AD) \leq 96128 ; \text{LM6}$$

$$\text{IF } (A1) > 244619 \text{ And } (AD) \leq 114221 \text{ And } (AD) > 96128 \text{ And } (AD) \leq 99978 ; \text{LM7}$$

IF (A1) > 244619 And (AD) <=114221 And (AD) >96128 And (AD)>99978 And (A1) <=430885; **LM8**

IF (A1) > 244619 And (AD) <=114221 And (AD) >96128 And (AD)>99978 And (A1) >430885; **LM9**

IF (A1) > 244619 And (AD) >114221; **LM10.**

نلاحظ مما سبق بأن معادلات التنبؤ تقيدنا بالتبؤ بحجم المبيعات المستقبلي لأسبوع ما من الأسابيع حيث أننا في حال أردننا في أسبوع ما استهدف مجال محدد من رقم المبيعات نقوم بناء على المعادلات بتحديد حجم النفقات الإعلانية الواجب إنفاقها ويتم تحديد رقم مبيعات مستهدف من كل مادة من المتغيرات المستقلة ويمكن أيضاًأخذ نسبة كل مادة من المواد بالنسبة للمبيعات الإجمالية حسب البيانات التاريخية وتحديد مجال النسب كحد أدنى وحد أعلى من كل مادة من المواد الداخلة في المعادلة.

حيث في حال اردننا زيادة إحدى المواد من أجل بلوغ الهدف نقوم بزيادة مبيعاتها بناء على علاقات الاقتران الناجمة معنا بحسب طرق العرض أو مختلف السياسات التسويقية الأخرى في قواعد الاقتران.

9.2.3 استخدام طريقة الانحدار المتعدد في التنبؤ:

يعتبر أسلوب تحليل الانحدار المتعدد من الطرق الإحصائية الجيدة في التنبؤ ويهدف إلى دراسة العلاقة البيانية بين متغير تابع وأكثر من متغير مستقل بواسطة معادلة رياضية وقد قام الباحث بعملية تحليل البيانات والتبؤ بالمبيعات بواسطة هذه الطريقة وذلك على مرحلتين

- تم تحليل المتغيرات المستقلة والتابعة التي قمنا بتحليلها عن طريق خوارزمية M5P والتأكد من أنها خطية.

- استخدام نموذج معادلات الانحدار عن طريق برنامج Weka لبناء نموذج التنبؤ.

أولاً: التأكد من أن المتغيرات خطية:

قام الباحث من أجل التأكد من أن المتغيرات خطية بتحليل هذه البيانات على البرنامج الاحصائي Spss، حيث قام بإدخال جميع المتغيرات المستقلة والمتغير التابع وتم تحليلها باستخدام الطريقة الاعتيادية (Enter) حيث ظهرت النتائج التالية:

الجدول رقم 8 ملخص تحليل النموذج

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.923 ^a	.853	.837	356790.2674	.853	53.919	13	121	.000

الجدول رقم 9 جدول تحليل التباين

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	8.923E+13	13	6.864E+12	53.919	.000 ^b
Residual	1.540E+13	121	1.273E+11		
Total	1.046E+14	134			

المصدر: مخرجات برنامج .Spss

الجدول رقم 10 لتحليل معنوية المتغيرات coefficient

Model	Unstandardized Coefficients					95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
	B	Std. Error	Standardized Coefficients Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1 (Constant)	643050.301	167011.370		3.850	.000	312407.442	973600.161					
A1	4.796	1.963	.793	2.443	.016	.909	8.682	.819	.217	.085	.012	86.640
A2	35.409	6.723	.926	5.279	.000	22.100	40.790	.013	.433	.184	.040	25.279
A3	7.120	10.845	.091	.657	.513	-14.351	28.592	.755	.060	.023	.064	15.664
A4	-2.897	3.770	-.268	-.768	.444	-10.361	4.668	.003	-.070	-.027	.010	99.707
A5	-15.323	5.083	-.970	-3.015	.003	-25.385	-5.260	.775	-.264	-.105	.012	85.134
A6	15.796	14.378	.206	1.099	.274	-12.669	44.261	.761	.099	.038	.035	28.761
A7	5.439	5.143	.259	1.057	.292	-4.744	15.621	.797	.096	.037	.021	47.156
A8	5.091	6.523	.140	.780	.437	-7.823	18.005	.837	.071	.027	.038	26.547
A9	0.425	9.514	.172	.006	.370	-10.411	27.260	.011	.000	.031	.032	31.160
A10	-45.295	9.080	-.712	-4.989	.000	-63.270	-27.319	.741	-.413	-.174	.060	16.733
A11	-19.364	18.022	-.147	-1.074	.285	-46.603	16.325	.787	-.097	-.037	.046	15.501
A12	1.421	5.366	.055	.265	.792	-9.203	12.045	.838	.024	.009	.028	35.658
AD	10.894	1.549	.413	7.026	.000	7.817	13.952	.791	.538	.245	.362	2.840

المصدر: مخرجات برنامج .Spss

نلاحظ من جدول تحليل التباين (Anova) جدول رقم (8) مailyi: ان قيمة $F=53.919$ من أجل $(F(13.121, df=0)$ وان قيمة $sig=0$ المرافق لـ الإحصائية الاختبار (F) مما يشير إلى معنوية معاملات الانحدار، وهذا يعني أن المتغيرات المستقلة مجتمعة لها تأثير معنوي في الانحدار أو أن واحدة على الأقل من معالم النموذج لها تأثير معنوي، لكن المهم بالنسبة لنا أن تكون جميع معالم النموذج معنوية، ولهذا السبب يتم اختبار معنوية جميع معالم النموذج كل على حدة حتى يتم قبوله.

ومن الجدول رقم (9) coefficient (A1,A2,A5,A10,AD) نلاحظ معنوية خمسة متغيرات مستقلة وهي $\alpha=0.05$. وعدم معنوية سبع متغيرات عند مستوى دلالة . $\alpha=0.05$

ونلاحظ من جدول رقم (7) ملخص تحليل النموذج ارتفاعاً في قيمة معامل الارتباط بين المتغير التابع والمتغيرات المستقلة وفي الوقت نفسه انخفاض معاملات الارتباط الجزئية بين المتغير التابع والمتغيرات المستقلة وهذه من مؤشرات الارتباط الخطي المتعدد وللحقيق من الأمر تم حساب معامل تضخم التباين VIF وهي تشير إلى تجاوز 12 متغير للقيمة 10 مما يؤكد أن هذه المتغيرات ترتبط بعضها بعلاقة خطية.
ثانياً: بناء نموذج معادلات الانحدار:

تم بناء النموذج باستخدام برنامج Weka حيث استخدمت خوارزمية الانحدار الخطي Linear Regression وقد ظهرت لنا النتائج التالية:

ملخص عملية تقييم النموذج	
معامل الارتباط	0.9202
متوسط الخطأ المطلق	269210
متوسط الجذر التربيعي للخطأ	346962
الخطأ المطلق النسبي	%38.7913
الجذر التربيعي للخطأ النسبي	%39.1426
عدد سجلات الاختبار	138

من خلال النتائج نلاحظ بأن نسبة الخطأ المطلق النسبي مرتفعة وهي / 38.7913 % بينما النسبة في تحليل أشجار القرار هي / 19.0283 % وهذا يدل على أن التنبؤ بالطرق الحديثة بواسطة أشجار القرار هي أفضل من الطرق الإحصائية التقليدية.

النتائج والتوصيات

Results and Advices

النتائج:

1. إمكانية بناء نموذج شجرة قرارات للتنبؤ بحجم المبيعات بالاعتماد على تحليل قواعد الاقتران، من خلال التقييب في بيانات الهايبر الماركت.
2. امكانية قواعد الاقتران من اكتشاف والاجابة عن كثير من التساؤلات المتعلقة بعلاقة المواد مع بعضها البعض وتبيين ذلك جلياً من نتائج تحليل قواعد الاقتران .
3. إمكانية الاستفادة من المواد المتولدة من تحليل قواعد الاقتران في عملية التنبؤ بالمبيعات وفي زيادة دقة هذا التنبؤ .
4. ان خوارزميات التنبؤ بالمبيعات باستخدام تقنيات التقييب في البيانات تعطي نتائج أدق في عمليات التنبؤ مقارنة مع الطرق الإحصائية التقليدية، وتبيين ذلك جلياً من خلال عملية المقارنة مع معادلات الانحدار حيث كان هناك فرق واضح في دقة التنبؤ.

التوصيات:

بناء على النتائج التي توصلت إليها الدراسة يوصي الباحث بما يلي:

1. اعتماد الهايبر ماركت على تحليل قواعد الاقتران من أجل العمل باحترافية ودقة في تلبية متطلبات الزبائن.
2. استخدام خوارزمية M5P من أجل بناء نموذج بسيط وسهل الفهم للتتبؤ بحجم المبيعات.
3. تطوير النموذج السابق ليشمل تحليل كافة أصناف المواد في الهايبر ماركت لاكتشاف مواد وعلاقات أخرى.
4. تطوير نموذج التتبؤ بإدخال متغيرات مستقلة خارجية تؤثر في حجم المبيعات مثل متغير التضخم والمتغيرات الاجتماعية والسياسية.
5. السعي إلى تطبيق النموذج وإدخال عامل الزمن في دراسة قواعد الاقتران لزيادة دقة التتبؤ والنتائج.



أ. المراجع العربية:

- الكتب:

1. أبو زيد كمال خليفة، مرعى عطية. 2004 - **مبادئ المحاسبة الإدارية الحديثة**، الدار الجامعية، الإسكندرية. مصر.
2. البكري صونيا. 2001 - **إدارة العمليات والإنتاج**، الدار الجامعية، الإسكندرية، مصر.
3. الجريسي خالد. 2010 - **سلوك المستهلك دراسة تحليلية للقرارات الشرائية للاسرة السعودية** ، مؤسسة الجريسي للتوزيع والاعلان، الرياض، السعودية.
4. الحسين محمد ابديوي.، 2012 - **تخطيط الإنتاج ومراقبته**، دار المناهج للنشر والتوزيع، عمان، الأردن.
5. دودي أحمد يوسف.، 2012 - **إدارة الإنتاج والعمليات**، عمان الأردن، دار الأكاديميون للنشر والتوزيع.
6. الدوري مؤيد عبد الرحمن.، حسين محمد سلامة.، 2013 - **أساسيات الإدارة المالية**، دار الراية للنشر والتوزيع، الأردن
7. رباعة علي - ذياب فتحي..، 1990 - **إدارة المبيعات**، منشورات مركز الخوارزمي، عمان.
8. الطويل هالة.، 2009 - **التقريب عن البيانات**، دار شعاع.
9. عبد الحميد طلعت اسعد.، 2000 - **دليل مدير المبيعات الفعال : كيف تدير العملية البيعية بكفاءة**، مؤسسة الأهرام، مصر.
10. عبد الحميد طلعت اسعد.، 2010 - **التسويق الفعال ، رؤية للنشر والتوزيع**، القاهرة.
11. عبيات محمد، الضمور هاني، حداد شفيق.، - 2005 ، **إدارة المبيعات والبيع الشخصي**، دار وائل للنشر، عمان الأردن.
12. عدون ناصر.، 2004 - **مراقبة التسيير في المؤسسة الاقتصادية**، دار المحمدية العامة، الجزائر.
13. فركوس محمد..، 2001 - **الموازنة التقديرية أداة فعالة للتسيير**، ديوان المطبوعات الجامعية، الجزائر.

14. كاظم أموري هادي..، 2009 - **مقدمة في القياس الاقتصادي**، زهران للنشر، عمان، الأردن.
15. الكرخي مجيد..، 2014 - **تخطيط وتقدير البرامج**، دار المناهج للنشر والتوزيع، عمان، الأردن.
16. مصطفى يوسف كافي..، 2012- **المحاسبة الإدارية**، مكتبة المجتمع العربي للنشر والتوزيع، الأردن.
17. موسى غانم فنجان..، 1999 - **إدارة المبيعات والاعلان**، دار الحكمة للطباعة، بغداد، العراق
18. نجم عبود نجم..، 2007- **مدخل الى إدارة العمليات**، دار المناهج للنشر والتوزيع، عمان.
19. الهواسي محمود..، البرزنجي حيدر..، 2014 - **مبادئ علم الإدارة الحديثة**، دار الكتب والوثائق العراقية، بغداد، العراق.
- **الرسائل الجامعية والدوريات:**
1. الجاعوني فريد..، 2008 - **أسلوب تحليل الانحدار الخطي المتعدد في دراسة أهم المتغيرات الاقتصادية والاجتماعية والديمografية المؤثرة في معدل الولادات الكلية**، مجلة جامعة دمشق للعلوم الاقتصادية والقانونية -المجلد - 24 العدد الثاني 2008.
 2. حامد جمال..، 2003 - **أساليب التنبؤ**، مجلة جسر التنمية، العدد الرابع عشر.
 3. رضا صباح، زياب السيد..، 2008 - **استخدام بعض الطرق الإحصائية والتصنيف الشجيري في التصنيف والتنبؤ بافلس الشركات مالياً**، مجلة العلوم الاقتصادية والإدارية، ع49، جامعة بغداد.
 4. رياض قادری..، 2011 - **طرق وأساليب التنبؤ عن المبيعات دراسة حالة**، رسالة ماجستير كلية علوم التسيير والاقتصاد، الجزائر
 5. الطائي فاصل عباس..، 2009 - **التنبؤ والتعميد للسلسل الزمنية باستخدام التحويلات مع التطبيق**، المؤتمر العلمي الثاني للرياضيات والاحصاء والمعلوماتية، جامعة الموصل.
 6. طعمه سعدية عبد الكريم..، 2012 - **استخدام تحليل السلسل الزمنية للتنبؤ بأعداد المصايبين بالأورام الخبيثة في محافظة الانبار**، مجلة جامعة الأنبار للعلوم الاقتصادية والإدارية، العدد 8.
 7. عبد الفتاح زكريا فريد..، 2004 - **اعداد الميزانية التخطيطية**، جامعة عين شمس.
 8. العبيد علي جاسم، ظاهر عبد القدير..، 2009 - **أهمية وتأثير دراسة السوق في رفع كفاءة قرارات الإنتاج في الوحدة الاقتصادية "دراسة ميدانية في الشركة العامة للصناعات الكهربائية /بغداد / " للفترة 1995-2006**، مجلة الادارة والاقتصاد، العدد الثامن والسبعين.

- .9. فاتح لقرقي..، 2014 - جودة نماذج السلسل الزمنية الموسمية المختلطة في SARIMA في التنبؤ بالمبيعات دراسة حالة مؤسسة مطاحن جديع بتقرت، رسالة ماجستير، كلية الاقتصاد، الجزائر.
10. المتى حسان..، 2009 - التنبؤ وتطبيقاته في ادارة الاعمال، رسالة ماجستير، جامعة دمشق.
11. مخرش عبلة..، 2006 - تقدير نموذج للتنبؤ بالمبيعات باستخدام السلسل الزمنية (نماذج بوكس وجينكينز) دراسة حالة الشركة الوطنية للكهرباء والغاز ، رسالة ماجستير ، كلية الحقوق والعلوم الاقتصادية، الجزائر.

المراجع الأجنبية:

- الكتب:

1. Agrawal, R., & Srekant, R., 1994 - **Fast Algorithm For Mining Association Rules**, Morgan Kaufmann Publishers, USA.
2. Berry M, Lionoff G., 2004 – **Data Mining Techniques For Marketing, Sales And Customer Relationship Management**, A John Wiley & Sons Inc, USA.
3. Brown M., 2014 - **Data Mining For Dummies**, a john wiley & sons inc, Canada.
4. DANIEL T.L., 2005 - **Discovering Knowledge In Data An Introduction To Data Mining**, Published by John Wiley & Sons Inc, USA.
5. Dolado J, Rodríguez D, Riquelme J, Cuadrado J, - 2007 - **A Two Stage Zone Regression Method For Global Characterization Of A Project Database**, Idea Group Inc, Canada.
6. Han, J. and Kamber,M. 2012- **Data Mining: Concepts And Techniques**, Morgan Kaufmann Publishers, USA.
7. Hancock M., 2012 – **Practical Data Mining**, taylor & Francis Group, Abingdon, UK.
8. K pujari.Arun., 2001 - **Data Mining Techniques**, published by Universities Press (india).
9. Kantardzic M., 2011 - **Data Mining: Concepts, Models, Methods, And Algorithms, 2nd Edition**, John Wiley & Sons.
10. Larose D., Larose C., 2005 - **Discovering Knowledge In Data An Introduction To Data Mining**, A John Wiley & Sons Inc, USA.

11. Mayatt G., Johnson W., 2014 - **Making Sense Of Data I. A Practical Guide To Exploratory Data Analysis And Data Mining.** 2nd Edition, John Wiley & Sons Inc, USA.
12. Tang, Z., & MacLennan, J., 2008 - **Data Mining With SQL Server 2008**, John Wiley & Sons Inc, USA.
13. Ville B., 2011 - **Decision Tree For Business Intelligence And Data Mining: Using SAS Enterprise Miner**, CRC Press, New York, USA.
14. Witten I.H., E. Frank, Hall M.A., 2011 - **Data Mining Practical Machine Learning Tools And Techniques.-3rd Ed**, Morgan Kaufmann Publishers, USA.

- الرسائل الجامعية والدوريات:

1. Alka G., Ravindra P., 2009 - **Building Privacy-Preserving C4.5 Decision Tree Classifier On Multiparties**, International Journal on Computer Science and Engineering Vol.1(3).
2. Azadi S, Amiri H, Rakhshandehroo R., 2016 - **Evaluating The Ability Of Artificial Neural Network And PCA-M5P Models In Predicting Leachate COD Load In Landfills**, ScienceDirect magazine, Vol55, p220-230.
3. Bhambri V., 2011 - **Application Of Data Mining In Banking Sector**, International Journal of Computer Science and Technology India, Vol 2.
4. Bharati M, Ramageri,. 2010 - **Data Mining Techniques And Applications**, Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305.
5. Carty D., 2011 - **An Analysis Of Boosted Regression Trees To Predict The Strength Properties Of Wood Composites**, master thesis, University of Tennessee, Knoxville.
6. Chi Huynh H., 2015 - **Improving M5 Model Tree By Evolutionary Algorithm**, Master's Thesis in Computer Science, Halden, Norway.
7. Deepa C, SathiyaKumari K, Pream Sudha V., 2010 - **Prediction Of The Compressive Strength Of High Performance Concrete Mix Using Tree Based Modeling**, International Journal of Computer Applications (0975 – 8887).
8. Frawley W., Piatetsky-Shapiro G., and Matheus C., 1992 - **Knowledge Discovery In Databases: An Overview**, AI Magazine, Vol 13, Number 3.

9. Gancheva, V., 2013 - **Market Basket Analysis Of Beauty Products**, Master of Science in Economics and Business, Erasmus University Rotterdam, Holland.
10. George Eapen A., 2004 - **Application Of Data Mining In Medical Applications**, master thesis, Systems Design Engineering, Ontario, Canada.
11. Ikonomovska E., 2012 - **Algorithms For Learning Regression Trees And Ensembles On Evolving Data Streams**, Doctoral Dissertation, Ljubljana, Slovenia.
12. Khalilzadeh N., 2008 - **Sales Prediction For Pharmaceutical Distribution Companies- A Data Mining Based Approach**, Master Thesis, Lulea university of technology, Sweden.
13. Kinski A., 2016 - **Google Trends As Complementary Tool For New Car Sales Forecasting: A Cross-Country Comparison Along The Customer Journey**, master thesis, university of twente, berlin.
14. Liao T., Triantaphyllou E., 2007 - **Recent Advances In Data Mining Of Enterprise Data Algorithms And Applications**, series on computers and Operations Research, vol 6, world scientific, London.
15. Madani S., 2009 - **Mining Changes In Customer Purchasing Behavior**, Master Thesis, Lulea university of technology, Sweden.
16. makridakis s., 1994 - **Forecasting : Its Role And Value For Planning And Strategy**, ScienceDirect magazine, vol, USA.
17. Onyari E, Ilunga F., 2013 - **Application Of MLP Neural Network And M5P Model Tree In Predicting Streamflow: A Case Study Of Luvuvhu Catchment**, South Africa, International Journal of Innovation, Management and Technology, Vol. 4, No. 1.
18. Padhy N., 2012 - **The Survey Of Data Mining Applications And Feature Scope**, International Journal of Computer Science, Engineering and Information Technolog, Vol.2, No.3.
19. Pal J., 2011- **Usefulness And Applications Of Data Mining In Extracting Information From Different Perspectives**, annals of library and information studies vol.58.
20. Pejic Bach M., 2003 - **Data Mining Applications In Public Organizations**, Information Technology Interfaces, 2003, Proceedings of the 25th International Conference on.
21. Rajcan S., 2013 – **Magazine Sales Prediction**, master thesis, Charles University in Prague

22. Ranjan J., 2007 - **Applications Of Data Mining Techniques In Pharmaceutical Industry**, Journal of Theoretical and Applied Information Technology, India.
23. Singh P, Agrawal S., 2013 - **Node Localization In Wireless Sensor Networks Using The M5P Tree And Smoreg Algorithms**, International Conference on Computational Intelligence and Communication Networks, Mathura, India.
24. Sondwale P., 2015 - **Overview Of Predictive And Descriptive Data Mining Techniques**, International Journal of Advanced Research in Computer Science and Software Engineering, Vol 5.
25. Svetina, M. & Zupančič, J., 2005 - **How To Increase Sales In Retail With Market Basket Analysis**, University of Maribor, Kranj, Slovenia
26. Thuraisingham B., 2000 - **A Primer For Understanding And Applying Data Mining**, IT Pro IEEE Xplore, Vol 2, pp 28-31.
27. Zhao Y., Chen Y., and Yao Y., (2006) - **User-Centered Interactive Data Mining**. Proceedings of the Sixth IEEE International Conference on Cognitive Informatics (ICCI'06): 457-466.
28. Zhu C, Xue M., 2009 - **The Application Of Data Mining In The Decision Of Supermarket Extension And Businesses Expansion Based On Evolutionary Computation**, ieee xplore, Circuits, Communications and Systems, 2009. PACCS '09. Pacific-Asia Conference on.

- مقالات منشورة على شبكة الانترنت:

<http://www.cs.waikato.ac.nz/ml/weka/>.

(١) **بيانات التنبؤ بالمبيعات**

بيانات التنبؤ بالمبيعات

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	ad	sales total
121509.1	28130.2 2	9262.43 4	66253.9 7	48461.1 8	8667.71 5	30052.7 2	24112.8 5	15152.0 9	9377.60 1	5716.60 8	35330.9 2	39804	160738 2
271746	62911.1 4	20714.7 4	148172. 1	108379. 8	19384.7	67210.6 6	53926.6	33886.5 3	20972.3 1	12784.7 7	79014.9 7	89018	359479 0
236521.5	54756.4 3	18029.6 5	128965. 5	94331.2 9	16871.9 9	58498.6	46936.4 7	29494.0 5	18253.8 1	11127.5 7	68772.8	77480	312882 2
208671.6	48308.9 5	15906.6 8	113780. 1	83223.9 9	14885.3 5	51610.5 2	41409.8	26021.1 9	16104.4 6	9817.31 9	60674.9 6	68356	276041 0
190523.8	44107.6 1	14523.3 1	103884. 9	75986.1 4	13590.8	47122.0 4	37808.4 6	23758.1 7	14703.8 9	8963.52 4	55398.1 6	62412	252034 2
213838.5	49505.1 2	16300.5 4	116597. 4	85284.6 8	15253.9 3	52888.4 4	42435.1 4	16503.2 2		62177.3 2		70049	282876 0
225609.6	52230.2 1	17197.8 4	123015. 7	89979.3 2	16093.6	55799.7 7	44771.0 6	28133.3 5	17411.6 7	10614.2	65599.9 8	73905	298447 4
241619.9	55936.7 1	18418.2 7	131745. 5	96364.6 7	17235.6 8	59759.5 8	47948.2 2	30129.8 2	18647.2 8	11367.4 3	70255.2 5	79150	319626 6
212805.8	49266.0 6	16221.8 3	116034. 4	84872.8 4	15180.2 6	52633.0 4	42230.2 2	26536.7 3	16423.5 3	10011.8 2	61877.0 6	69711	281510 0

214410.6	49637.5 6	16344.1 6	116909. 4	85512.8 5	15294.7 4	53029.9 3	42548.6 7	26736.8 4	16547.3 8	10087.3 2	62343.6 7	70236	283632 8
215666.1	49928.2 3	16439.8 6	117594	86013.5 9	15384.3	53340.4 7	42797.8 3	26893.4	16644.2 7	10146.3 9	62708.7 4	70648	285293 7
241214.1	55842.7 7	18387.3 4	131524. 2	96202.8 3	17206.7 3	59659.2 2	47867.6 9	30079.2 2	18615.9 7	11348.3 4	70137.2 6	79017	319089 8
218695	50629.4 5	16670.7 5	119245. 5	87221.6 1	15600.3 6	54089.6 1	43398.9	27271.1 1	16878.0 3	10288.8 9	63589.4 5	71640	289300 5
230897.8	53454.4 7	17600.9 5	125899. 2	92088.4	16470.8 3	57107.7	45820.4 7	28792.7 8	17819.8	10862.9 9	67137.6 2	75637	305442 9
278118.3	64386.3 5	21200.4 9	151646. 6	110921. 2	19839.2 5	68786.7	55191.1 4	34681.1 4	21464.0 9	13084.5 6	80867.8 1	91106	367908 5
484329.1	112125. 6	36919.5 9	264085	193163. 8	34549.0 7	119788. 6	96112.6 1	60395.4 8	37378.6 4	22786.1 1	140827. 3	15865 6	640694 3
344306.1	79709.3	26245.8 7	187736. 1	137318. 8	24560.6 9	85156.8 6	68325.7 7	42934.7 2	26572.2 1	16198.4 8	100113. 1	11278 8	455465 0
339406.9	78575.1 2	25872.4 8	185064. 8	135364. 1	24211.2 5	83945.1 5	67353.5 5	42323.7 9	26194.1 1	15967.9 9	98688.5 7	11118 3	448984 1
320245.1	74139.0 1	24411.7 4	174616. 6	127722. 6	22844.3 2	79205.8 7	63550.9 9	39934.3 2	24715.2 7	15066.4 9	93116.9 3	10490 6	423635 9
307484.3	71184.7 9	23439.0 1	167658. 7	122633. 2	21934.0 5	76049.7 6	61018.6 7	38343.0 6	23730.4 5	14466.1 3	89406.5	10072 5	406755 3
302923.1	70128.8 6	23091.3 2	165171. 7	120814. 1	21608.6 8	74921.6 6	60113.5 4	37774.2 9	23378.4 3	14251.5 5	88080.2 7	99231	400721 6
296936.6	68742.9 3	22634.9 8	161907. 5	118426. 5	21181.6 4	73441.0 1	58925.5 4	37027.7 7	22916.4 2	13969.9	86339.5 7	97270	392802 3
321999.9	74545.2 7	24545.5 1	175573. 5	128422. 5	22969.5	79639.9	63899.2 3	40153.1 5	24850.7 1	15149.0 5	93627.1 8	10548 0	425957 3
295088.7	68315.1 3	22494.1 1	160899. 9	117689. 5	21049.8 2	72983.9 7	58558.8 3	36797.3 4	22773.8	13882.9 6	85802.2 6	96665	390357 8
331176.6	76669.7 2	25245.0 3	180577. 1	132082. 4	23624.1 1	81909.5 5	65720.2 8	41297.4 7	25558.9 2	15580.7 8	96295.4 5	10848 7	438096 6

296368.9	68611.5	22591.7	161597.9	118200.1	21141.1	73300.6	58812.8	36956.9	22872.6	13943.1	86174.5	97084	392051
313959.9	72683.9	23932.6	171189.3	125215.5	22395.9	77651.3	62303.7	39150.5	24230.2	14770.7	91289.3	10284	415321
308537.4	71428.5	23519.2	168232.9	123053.9	22009.1	76310.2	61227.6	38474.3	23811.7	14515.6	89712.7	10107	408148
325917.8	75452.2	24844.1	177709.7	129985	23248.9	80608.9	64676.7	40641.7	25153.0	15333.3	94766.3	10676	431140
291814.5	67557.1	22244.5	159114.6	116383.7	20816.2	72174.1	57909.0	36389.0	22521.1	13728.9	84850.2	95592	386026
301601.6	69822.9	22990.5	164451.1	120287	21514.4	74594.8	59851.2	37609.5	23276.4	14189.3	87696.0	98798	398973
298271	69051.8	22736.6	162635	118958.7	21276.8	73771.0	59190.3	37194.1	23019.4	14032.6	86727.5	97707	394567
264294.3	61186.0	20146.7	144109	105407.9	18853.1	65367.6	52447.8	32957.3	20397.2	12434.1	76848.2	86577	349621
587959	138610	56850	354155	215456	42384	190788	98432	67478	63852	29208	166442.	13429	478640
559916.2	130978	36010	328258	233896	36598	158148	107176	61196	60020	28286	164045.	11409	406632
517465.3	123378	48940	327652	215670	39646	158464	94840	77800	65810	29570	150202	11147	397293
542229.7	118190	48200	356361	224748	44366	170976	95680	76512	62170	20402	145114	11526	410806
604273	114296	39250	333848	235122	48384	166428	110544	64762	52376	26740	149040	11480	409173
636544.5	107684	40920	335209	230606	39586	152342	115066	68782	62814	27340	138902	11435	407552
601982.9	103430	45920	331273	226062	45848	181542	107102	78926	53906	32658	153082	11756	419009
563109.1	99196	47630	348518	232398	45952	140424	102972	81698	50010	31458	145988	11487	409408
												1	0

633934.6	101102	41300	336608.3	251454	56568	168380	101426	76388	47721.62	29582	145626	126917	4523393
576688.4	101982	37080	282510	231636	45786	129636	90120	72166	44585.5	27690	133558	111792	3984348
575128.9	102284	43650	304070.6	228722	47510	132414	112631.9	72906	52807.42	25584	148770	113662	4051007
444722.1	80123	38790	226788.3	180580	39256	101525	80791	54255	33748	15786	127711	94401	3364521
656384.5	131080	46840	357939.5	279792	47791	163302	123369	76525	45764.5	31566	162549	101575	3620211
261007.8	48496	22890	134278	101973	19811	60441	52720	33909	18716.8	13345	77208.75	106147	3783144
218290.8	42622	23015	109805	93703	15864	50004	47379	28029	13944	9771	62236	89429	3187300
199826.6	41559	15167	97917	76420	11811	35047	38048	25431	11895	6454	60544	77463	2760841
185644.3	41911	18360	100647	80326	13866	44833	43710	25663	14490.5	6694	63102	83986	2993325
203868.6	50138	18964	110015	80691	16211	53861	43275	31490	13277	8128	63211	88593	3157514
200637.8	50077	15421	107937	84505	13945	49872	44089	25424	14099	7608	65587	89515	3190361
184847.4	48554	15999	105280	76294	10299	37022	50764	23237	10273	8608	66773	84761	3020922
190991.8	47721	14067	107993	73379	9773	43359	46856	24598	17915	10851	62687	80215	2858923
185893.2	53453	16218	137970	79312	14823	51700	55427.68	28883	13131.9	9778	70823	91177	3249623
179236.8	46510	15387	94020	74198	11810	47987	49150	28737	10734	13496	70497	84357	3006544
175931	43622	12702	84146.5	69691	8842	38783	43249	26192	6688	10278	62934	76242	2717328

161600.7	47627	10670	92109.5	65388	9336	34730	44665.6 8	27645	11098.5	8424	62401	78481	279709 9
190619.2	49498	11616	87123	67649	7803	32347	42325	34195	11857.4	7685	61615	79609	283731 0
182137.1	47055	13795	87205	67480	10156	36097	45091	29507	9800	6184	68945	78695	280473 3
181170.5	49004	14001	105067	67365	8017	42022	50038	23520	7019.35	7402	72600	80058	285333 6
194115.4	45614	20692	93670	64004	9746	32779	42199	20641	6564	6758	69027	76204	271596 2
150742.9	44363	14154	88476	60621.6 5	7537	27327	43827	20517	7875	6870	64253	71993	256587 2
204790.6	48355	8871	104963	61305	7578	34171	44087	29909	7760	3406	54657	25517	274681 7
172275.3	39208	15726	72913	52843	10463	26973	33861	18893	6775	3074	57003	16595	237578 7
168168	46282	15082	67429.2 5	58220	9530	31778	40113	23070	11345	5169	63821	22511	294791 9
247619.2	60109	8722	103890	57598	12860	61707	53265	20992	17135	9309	80498	34826	381963 6
193396.2	56229	8791	82960	52487	13626	42404.7 2	52718	15470	15171	6955	84376	35133	322033 3
189678	56420	12362	91061.1 9	57461	12392	51831.7 5	48143	17503	9417	7040	68486	41026	352724 5
163701.3	55003	7860	84768.1 9	56942	10112	54001.3 8	44019	12704	7913	8102	54383	36302	311728 3
140730.7	41962	6335	80423	43218	8849	28077.7 2	27972	17200	11738	6007	51539	27711	233250 1
178125.7	49326	11430	92341	70274	9189	28540.1 1	31390	19692	12947	7747	46612	37113	275526 4
184138.6	46359	7540	94785.0 6	71311	10670	36419.6	34338	15637	8641	8385	50354	35617	284180 2

154135.8	45093	11259	90174.0 8	75358	11084	40098.4 8	39323	24686	8809	8041	55521.9 6	36361	284559 4
172258.9	50828	15366	94019	72951	9737	36808.2 8	37533	16633	7125	7872	46484	36486	276113 4
161362.5	53174	10988	76624	72883	9712	36636.8 1	29122	17659	9983	6496	54075	36789	267375 0
176817.9	53765	8629	82027	77774	9573	45526.7	25884	18223	7896	11100	44682	39971	282475 4
158315	41693	8597	87945	76962	17781	38665	25861	18590	6185	10520	44587	34374	252175 5
141717.4	46196	13119	77235	73059	9863	38072	22217	11659	10413	6412	33463	33934	237698 9
129632.2	46481	11380	67237	69439	12762	37428.7 2	19110	17032	3685	9002	32084	38738	277777 6
106520.6	44035	8605	51459	53369	9058	22927.3 6	24541	8230	8045	6731	29724	30448	183371 9
130206.5	50155	12765	86540	75111	8721	29337	21677	17918	9006	10039	39076	36493	229357 4
129814	43352	14235	80317	69037	10407	28087	19495	18335	6824.5	7980	43256	34280	219881 2
121636.5	40305	10680	75116	64556	12469	22370	18758	16751	12923.0 1	8374	29435	31856	212036 8
114679.6	42967	16205	78837	66599	13359	29053	18716	17733	7886	10174	34498	38261	235804 6
96393.3	33316	10300	64347	57251	9513	27221	18904	11442	5239	5006	28012	65823	217951 2
137171.5	41710	15720	72224.5	68831	8026	25495	21690	21868	4840	10320	30739	86518	225270 0
132071.1	36767	13230	62160	62807	11249	22715	23073	16599	10038	4155	28282	62752	212388 1
124116.2	35204	15650	67041	67923	10638	19257	20668	14223	5704	4836	26786	68778	212083 6

122306.2	31086	12949	77596	67687	10607	16815	23933	15879	6838		26982	71864	222787 5
118217.8	34398	13535	76923	63522	10452	14026	18576.5	11329	3529	2320	33736	70167	208400 7
120867.2	37922	10469	57376	58540	10992	12220	21338.5	14572	4731.38	950	31981	66798	210733 0
106087.4	33810	6142	59353	55380	10678	10511	27285	11341	4846	7061	30315	67689	194633 0
112220.6	34460	7776	52255	56481	9946	14124	22831	12677	5155	1980	38760	47596	197781 4
103317.8	32099	7479	57220	55774	9294	13529	28136	11663	3273	2996	41674	44950	188375 0
93513.76	33981	8001	61518	50108	8410	13009	23346	13706	5764	6205	41505	48909	192333 9
109121.7	29170	6437	63104	46629	7265	8502	23596	10026	4503	4260	37348	48897	196980 3
111432.1	33152	7258	58088	57942	8504	18568	29787.2 5	11114	5455	7016	32661	45364	211187 9
101489	31782	9183	63363	52105	12305	14275	35419.3 8	11179	3306	6069	34165	42918	205657 2
112600.6	33766	10832	66378	57494	9865	10594	34136	11823	8558	7356	22729	45744	207708 2
124036.4	36425	10500	99232	53103	11298	14233	32860.6 3	17234	12411	5410	23646	51392	261890 5
84008.1	33476	5622	48428	45625	9415	8995	33299.1 3	10632	8020	5047	32284	75420	210467 4
118326.3	38855	5626	75859	49316	14372	12744	32571	12479	9035.04	3648	35834	92305	282651 5
90191.23	33612	4916	68409	44427	11476	14136	44665	12427	6203	5073	36143	72778	212383 5
119825	40462	4617	73746	47578	10953	14375	24274	17541	7136	7281	33923	84288	262227 1

104097.9	33732	4076	51026	44985	8018	14419	25143	16217	10233	6764	37736	79629	220212 4
143602.3	41897	3691	68414	47109	10210	21928	26401	22779	15158	4591	42661	10147 3	291497 2
126386.6	40647	4693	61968	49499	8980	22378	22192	18154	11026	9157	45113	85691	241935 6
131599.6	43648	4166	67474	45414	7590	16583	27396	18300	11978	10012	44946	90361	269149 0
101890.1	41977	5688	63017	40303	9079	14869	30509	17613	6205	8052	52720	71612	226938 0
119537.3	39418	3458	65470	39961	7703	27981	24424	18684	10886.7 9	6890	42669	87163	253213 8
119110.9	44454	5912	59211	39405	6560	30699	34371	22062	14242	5412	49532	84168	257163 0
96991.58	40885	4317	50664	38361	4839	18881	38427.5	15233	12057	6778	46136	74012	238108 7
94932.57	44216	6850	53642	36451	5623	21111	29031.5	15720	10867	5575	35456	76983	223298 1
94416.31	39608	3458	57355	37849	6792	18514	28515	13223	10275	6605	38325	86770	228451 9
100277.9	43821	2476	59406	35194	5788	19390	29521	13644	11896	5305	52218	86439	231634 9
115239.5	58403	5737	60193.5	37228	6877	36318	42371.5	15043	17973	8588	60874	65373	326359 4
84195.46	51929	5621	56420	32858	5799	16360	42055.5	13204	10678	7136	58036.9 9	65853	251554 5
92551.49	54292	12306	63338	37015	1591	26656	38463.5	11885	18153	6338	79492	73876	279253 1
92352.75	53965	10378	49500	35235	1070	15986	32934	18688	10996	6690	57198	70238	251036 1
95329.43	52483	7507	53723	35480	3683	29158	41694	14554	16499	6578	59136	75683	297966 4

62023.43	37733		31968	27215	2038	14665	25006	13394	11455	2908	51690	51815	168260 9
64195.85	50796	2022	51447.5 5	43235	2395	14828	32490	17332	12793	5192	59897.8 9	14279	217549 2
56531.67	41257	9218	52643.5	41755	6935	15199	28038.5	12046	14251	5483	51993.6 3	15217	227107 1
67848.01	48571	13110	50386	42460	4389	24935	31337.5	14495	10512	6443	56511.1 3	16695	257894 5
56334.24	49521	6955	45730	42240	5334	26325	23038	12206	14824	5512	47826.8 8	15669	216685 2
51927.21	47587	9930	40469	41185	3784	28847	24310.5	11146	10836	6172	36190.2 5	17068	208997 9
64560.3	51287	7975	40195	43285	2268	20062	21458.5	15866	14243	6085	39709.1 4	16525	205084 6
55298.57	46925	10375	38013.8	38320	1528	27381	24715	7498	7557	6749	40149.2 6	16452	202407 1
57249.77	44564	14415	32192	39400	696	26871	16285	2308	9060	6333	35544	13851	185216 6
55391.87	42066	7440	31165	31965	606	23166	19479.5	4941	8727	5238	36841	6571	159179 7
53412.98	43757	8305	27576	33945		20387	17486	2655	6786	5217	34555	9031	150491 1
58513.1	45500	8755	33228	37630	2860	27376	16433	5813	4629	5964	35879	8683	159560 2
55927.17	37249	10672	25066	32790	2749	20739	21003.7 8	4964	4683	5312	34113.8 8	7582	151610 1
55856.26	39944	8304	34194	36480	8943	22918	25385.9 2	3284	6640.8	5701	34276.0 1	0	161334 6
69981.23	36943	6744	30216	34015	7170	16120	15462.5	3340	6883.1	5834	31812.7 6	0	158336 2
55664.86	35128	7970	25410	32860	4350	14826	11583.7 5	4500	5176	4960	41140.7 6	0	142538 9

32777.2	14534	4340	18210	16739	2176	8907	6512	5524	3630	5126	16471	0	205749
---------	-------	------	-------	-------	------	------	------	------	------	------	-------	---	---------------

/ ٢ /

بيانات التحليل

====Run information====

Scheme: weka.classifiers.trees.M5P -M 4.0

Relation: all

Instances: 138

Attributes: 14

A1

A2

A3

A4

A5

A6

A7

A8

A9

A10

A11

A12

ad

sales total

Test mode: split 75.0% train, remainder test

====Classifier model) (full training set)==== (

M5 pruned model tree:

)using smoothed linear models(

A1 <= 244619.546 :

| A7 <= 30375.858 : LM1 (61/17.368%(

| A7 > 30375.858 :

| | A8 <= 42348.25 :

| | | A2 <= 46239 : LM2 (7/10.285%(

| | | A2 > 46239 : LM3 (9/7.994%(

| | A8 > 42348.25 :

```
| | |A2 <= 50002.616 : LM4 (13/7.99%()
| | |A2 > 50002.616 : LM5 (13/13.739%()
A1 > 244619.546 :
|ad <= 114221 :
| | ad <= 96128.5 : LM6 (6/15.567%()
| | ad > 96128.5 :
| | |ad <= 99978 : LM7 (6/4.215%()
| | |ad > 99978 :
| | | |A1 <= 430885.689 : LM8 (10/0%()
| | | |A1 > 430885.689 : LM9 (5/0.001%()
|ad > 114221 : LM10 (8/13.044%()
```

LM num: 1

sales total =

```
* 1.4498A1
* 27.4905 +A2
* 1.5909 -A3
* 7.2851 +A4
* 1.8666 -A5
* 0.6924 -A7
* 8.0297 +A8
* 6.5094 +A10
* 5.6573 +ad
61807.6579 +
```

LM num: 2

sales total =

```
* 1.6316A1
* 24.5328 +A2
* 2.606 +A3
* 0.4292 -A4
* 1.8666 -A5
* 0.9232 -A7
* 11.1118 +A8
* 5.444 -A10
* 2.3559 +ad
985996.0976 +
```

LM num: 3

sales total =

* 1.6316A1
* 24.2471 +A2
* 2.606 +A3
* 0.4292 -A4
* 1.8666 -A5
* 0.9232 -A7
* 9.0787 +A8
* 5.444 -A10
* 2.3559 +ad
1090574.9731 +

LM num: 4

sales total =

* 1.6316A1
* 28.642 +A2
* 17.6565 +A3
* 1.9779 -A4
* 1.8666 -A5
* 0.9232 -A7
* 5.9624 +A8
* 5.444 -A10
* 2.3559 +ad
984109.0524 +

LM num: 5

sales total =

* 1.6316A1
* 34.5845 +A2
* 11.7982 +A3
* 1.9779 -A4
* 1.8666 -A5
* 0.9232 -A7
* 5.9624 +A8
* 5.444 -A10
* 2.3559 +ad
804708.1767 +

LM num: 6

sales total =

* 1.1243A1

* 10.9063 +A2
* 4.4051 -A5
* 12.8478 -A10
* 1.4351 +A12
* 12.9932 +ad
2185634.0192 +

LM num: 7

sales total =

* 1.1641A1
* 10.9063 +A2
* 4.4051 -A5
* 12.8478 -A10
* 1.2813 +A12
* 17.8148 +ad
1749251.2107 +

LM num: 8

sales total =

* 1.5508A1
* 10.9063 +A2
* 4.4051 -A5
* 12.8478 -A10
* 1.2813 +A12
* 19.2946 +ad
1490653.852 +

LM num: 9

sales total =

* 1.1028A1
* 10.9063 +A2
* 4.4051 -A5
* 12.8478 -A10
* 1.2813 +A12
* 20.2589 +ad
1515348.6614 +

LM num: 10

sales total =

* 1.6644A1

* 10.9063 +A2
* 4.4051 -A5
* 12.8478 -A10
* 26.6877 +ad
662426.8476 +

Number of Rules : 10

Time taken to build model: 0.17 seconds

====Evaluation on test split=====

Time taken to test model on training split: 0 seconds

====Summary=====

Correlation coefficient	0.9732
Mean absolute error	178600.6339
Root mean squared error	263775.2899
Relative absolute error	22.3352%
Root relative squared error	25.7852%
Total Number of Instances	34

====Run information=====

Scheme: weka.classifiers.trees.M5P -M 4.0

Relation: all

Instances: 138

Attributes: 14

A1

A2

A3

A4

A5

A6

A7

A8

A9
A10
A11
A12
ad
sales total

Test mode: evaluate on training data

====Classifier model) (full training set)==== (

M5 pruned model tree:
)using smoothed linear models(

A1 <= 244619.546 :

|A7 <= 30375.858 : LM1 (61/17.368%(
|A7 > 30375.858 :
| | A8 <= 42348.25 :
| | | A2 <= 46239 : LM2 (7/10.285%(
| | | A2 > 46239 : LM3 (9/7.994%(
| | A8 > 42348.25 :
| | | A2 <= 50002.616 : LM4 (13/7.99%(
| | | A2 > 50002.616 : LM5 (13/13.739%(
A1 > 244619.546 :

|ad <= 114221 :
| | ad <= 96128.5 : LM6 (6/15.567%(
| | ad > 96128.5 :
| | | ad <= 99978 : LM7 (6/4.215%(
| | | ad > 99978 :
| | | | A1 <= 430885.689 : LM8 (10/0%(
| | | | A1 > 430885.689 : LM9 (5/0.001%(
|ad > 114221 : LM10 (8/13.044%(

LM num: 1

sales total =

* 1.4498A1
* 27.4905 +A2
* 1.5909 -A3
* 7.2851 +A4
* 1.8666 -A5
* 0.6924 -A7

* 8.0297 +A8
* 6.5094 +A10
* 5.6573 +ad
61807.6579 +

LM num: 2

sales total =

* 1.6316A1
* 24.5328 +A2
* 2.606 +A3
* 0.4292 -A4
* 1.8666 -A5
* 0.9232 -A7
* 11.1118 +A8
* 5.444 -A10
* 2.3559 +ad
985996.0976 +

LM num: 3

sales total =

* 1.6316A1
* 24.2471 +A2
* 2.606 +A3
* 0.4292 -A4
* 1.8666 -A5
* 0.9232 -A7
* 9.0787 +A8
* 5.444 -A10
* 2.3559 +ad
1090574.9731 +

LM num: 4

sales total =

* 1.6316A1
* 28.642 +A2
* 17.6565 +A3
* 1.9779 -A4
* 1.8666 -A5
* 0.9232 -A7
* 5.9624 +A8

* 5.444 -A10
* 2.3559 +ad
984109.0524 +

LM num: 5

sales total =

* 1.6316A1
* 34.5845 +A2
* 11.7982 +A3
* 1.9779 -A4
* 1.8666 -A5
* 0.9232 -A7
* 5.9624 +A8
* 5.444 -A10
* 2.3559 +ad
804708.1767 +

LM num: 6

sales total =

* 1.1243A1
* 10.9063 +A2
* 4.4051 -A5
* 12.8478 -A10
* 1.4351 +A12
* 12.9932 +ad
2185634.0192 +

LM num: 7

sales total =

* 1.1641A1
* 10.9063 +A2
* 4.4051 -A5
* 12.8478 -A10
* 1.2813 +A12
* 17.8148 +ad
1749251.2107 +

LM num: 8

sales total =

* 1.5508A1

* 10.9063 +A2
* 4.4051 -A5
* 12.8478 -A10
* 1.2813 +A12
* 19.2946 +ad
1490653.852 +

LM num: 9

sales total =

* 1.1028A1
* 10.9063 +A2
* 4.4051 -A5
* 12.8478 -A10
* 1.2813 +A12
* 20.2589 +ad
1515348.6614 +

LM num: 10

sales total =

* 1.6644A1
* 10.9063 +A2
* 4.4051 -A5
* 12.8478 -A10
* 26.6877 +ad
662426.8476 +

Number of Rules : 10

Time taken to build model: 0.08 seconds

====Evaluation on training set=====

Time taken to test model on training data: 0.02 seconds

====Summary=====

Correlation coefficient	0.9807
Mean absolute error	132055.4112
Root mean squared error	175984.1378
Relative absolute error	19.0283%

Root relative squared error	19.8536%
Total Number of Instances	138

====Run information====

Scheme: weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8

Relation: all

Instances: 138

Attributes: 14

A1

A2

A3

A4

A5

A6

A7

A8

A9

A10

A11

A12

ad

sales total

Test mode: evaluate on training data

====Classifier model) (full training set)==== (

Linear Regression Model

sales total=

- * 7.1135 A1+
- * 36.3542 A2+
- * 14.6836- A5+
- * 42.8259- A10+
- * 11.2044 ad+

621349.5206

Time taken to build model: 0.02 seconds

====Evaluation on training set====

Time taken to test model on training data: 0 seconds

====Summary====

Correlation coefficient	0.9202
Mean absolute error	269210.1312
Root mean squared error	346962.9968
Relative absolute error	38.7913%
Root relative squared error	39.1426%
Total Number of Instances	138

Abstract

The research aimed to study consumer behavior in hypermarkets by analyzing the shopping basket using the algorithm of association rules through SQL server business intelligence.

From this analysis we discovered the rules and relationships between the items so we could build a model to predict the sales volume after adding additional independent variables and analyzing this historical data over three years using M5P algorithm via WEKA tool, and it has been using the percentage split feature at the data split process to training data and testing data to build the decision tree that support the prediction process.

The prediction of sales volume between decision trees and traditional statistical methods (regression equations) was also compare to reach the results that support the validity of the research problem.

Keywords; sales forecast, shopping basket analysis, association rules, trees, M5P.

DECLARATION

It is hereby declared that this work "**Building a Decision Tree to Predict the Volume of Sales by Using Data Mining Techniques (An Applied Study on a Hypermarket)**" has not already been accepted for any degree, and it is not being submitted concurrently for any other degree.

Candidate

Ahmad Mohammad Wahby Akkad

Certification

It is hereby certified that the work presented in this thesis is result of a research carried out by the candidate Ahmad Mohammad Wahby Akkad under the supervision of Dr. Ghassan Saket, Professor at Department of Statistics and Information Systems, Faculty of Economics University of Aleppo, and any other references mentioned in this work are documented in the text of this thesis.

Candidate

Ahmad Mohammad Wahby Akkad

Supervisor

Dr. Ghassan Saket



**Building a Decision Tree to Predict the Volume of Sales by Using
Data Mining Techniques**
(An Applied Study on a Hypermarket)

Presented To Obtain the Degree of Master at Managements Information Systems.

Submitted by
Ahmad Mohammad Wahby Akkad

Supervised by
Dr. Ghassan Saket

Professor at Department of
Statistics and Information Systems, Faculty of Economics University of
Aleppo

1439 AH - 2017 AD

**University of Aleppo
Faculty of Economics
Dept. of Statistics and Information Systems**



**Building a Decision Tree to Predict the Volume of Sales by Using
Data Mining Techniques**

(An Applied Study on a Hypermarket)

Presented To Obtain the Degree of Master at Managements Information Systems

**Submitted by
Ahmad Mohammad Wahby Akkad**

1439 AH - 2017 AD