Syrian Arab Republic
Ministry of Higher Education
Syrian Virtual University

الجمهورية العربية السورية
وزارة التعليم العالي
الجامعة الافتراضية السورية

تصميم نموذج نظام دعم القرار لإدارة الموارد البشرية

بالاعتماد على تقنيات الذكاء الصنعي

Design An Intelligent Decision Support System Prototype for Human Resources Management

بحث مقدم لنيل درجة ماجستير الدراسات العليا في إدارة التقانة

A Thesis submitted to the Syrian Virtual University in Partial Fulfillment of the Requirements for the M.A. Degree in Technology Management

إعداد م. جلال الضاهر
Submitted by Jalal AL-Daher
(Jalal-42398)

الأستاذ المشرف الدكتور طاهر رجب قدار
Supervised by Professor Taher Kaddar

2013-2014

الجمهورية العربية السورية
وزارة التعليم العالي
الجامعة الافتراضية السورية

ماجستير الدراسات العليا
إدارة التقانة
Master of
Technology Management

دمشق في _____

اسم الباحث: جلال الضاهر

عنوان الرسالة: تصميم نموذج نظام دعم القرار لإدارة الموارد البشرية بالاعتماد على تقنيات الذكاء الصنعي

قدمت هذه الرسالة لنيل درجة ماجستير الدراسات العليا في إدارة التقانة من الجامعة الافتراضية السورية

لجنة المناقشة والحكم:

تاريخ المناقشة:

موافقة مدير البرنامج	أجيزت الرسالة بتاريخ
2014/...../.....	...../...../2014
موافقة مجلس الشؤون العلمية	موافقة مجلس الجامعة
2014/...../.....	...../...../2014

الجمهورية العربية السورية
وزارة التعليم العالي
الجامعة الافتراضية السورية

ماجستير الدراسات العليا
إدارة التقانة
Master of
Technology Management

عنوان الرســـالة

تصميم نموذج نظام دعم القرار لإدارة الموارد البشرية

بالاعتماد على تقنيات الذكاء الصنعي

Design An Intelligent Decision Support System Prototype

for

Human Resources Management

الشـــــكر

بسم الله الرحمن الرحيم ...

﴿وما أوتيتـم من العلـم إلا قليلاً﴾

... صدق الله العظيم

الحمد لله رب العالمين الذي بحمده تدوم النعم ...

أتقدم بخالص الشكر والتقدير لكل من مد يد العون سواء برأي أو توجيه أو دعم وأخص بالذكر كلاً من الأستاذ المشرف الدكتور م. طاهر رجب قدار ومدير برنامج إدارة التقانة الدكتور م. ابراهيم شعيب المحترمين لدورهما الكبير في إنجاز هذا البحث ...

أتقدم بكامل الشكر والامتنان لعائلتي وزملائي الذين ساهموا بكامل إمكاناتهم لإتمام هذا العمل ...

والحمد لله رب العالمين ...

فهرس العنـــــــاوين

الملخص ..... ز

مقـــدمة ..... ز

الهدف من البحث ..... ز

أســـــلوب البحث ..... س

نتـــــائج البحث ..... س

قيمـــــــة البحث ..... س

الكلمات المفتاحية ..... ش

مقـــــــــــــــدمة ..... ص

مشـــــــــــكلة البحث ..... ض

الدراســـــــات المرجعية ..... ط

الدراســات المرجعية باللغة العربية ..... ط

الدراسات المرجعية باللغة الأجنبية ..... ق

ملخــص الدراســـات المرجعيـــة ..... ب ب

المقارنة بين الدراسة الحالية والدراسات المرجعية ..... ب ب

أهميـــــــة البحث ..... ث ث

هـــــــدف البحث ..... ث ث

فرضيـــات البحث ..... ج ج

منهـــــج البحث ..... ج ج

الفصل الأول: التمهيد النظري 1

1.1. تقييم أداء العـاملين 2

1.1.1. نبذة عن تقييم أداء العاملين في سورية 2

2.1.1. أهداف التقييـــــــم 3

3.1.1. نوع التقييم المستخدم 3

3.1.1.أ. تفصيلات التقييم نصف السنوي 4

3.1.1.ب. نقـــاط التقييـــم 5

4.1.1. نماذج تقييم الأداء المعتمدة 6
5.1.1. ملخص عملية تقييم الأداء في القطر العربي السوري 8
2.1. عملية التفقد اليومي .....
1.2.1. مقدمـــــة 9
2.2.1. عملية التفقد اليومي في الجمهورية العربية السورية 10
3.1. التنقيب في المعطيات .....
1.3.1. مقدمـــــــة 13
2.3.1. خوارزميات التصنيف 14
3.3.1. خوارزميات العنقـــدة 26
4.3.1. انتقـــــاء المعايير 31
4.1. تطـــبيق WEKA .....
1.4.1. مقدمـــــــــة 33
2.4.1. الميزات الرئيسية 34
3.4.1. المصطلحات الأساسية 35
3.4.1.أ. مصفوفة الشـــك 35
3.4.1.ب. معامل الإحصاء كابا (Kappa Static) 36
4.4.1. الحزمة البرمجية 37
5.4.1. الميـزات والفوائد 37
6.4.1. نقـــاط الضعف 38

الفصل الثاني: المنهج العملي 39
1.2. عينة البحث 40
1.1.2. النموذج الموحد لتقييم الأداء 40
2.1.2. تجهـيز معطيـــات البحث 42
2.2. التنقيب في البيانات 45
1.2.2 بنـاء نمـــوذج العمل 45
2.2.2. نتائج المنهج رقم /1/ 47
2.2.2. نتائج المنهج رقم /2/ 49

3.2. انتقـــــاء المعـــــــــايير 51
4.2. معالجة بيانات التفقد اليومي 52

الفصل الثالث: المناقشة 54
1.3. نتائج المنهج رقم /1/ 55
1.1.3. دقة التصنيف بالنسبة لبيانات التدريب 55
2.1.3. دقة التصنيف بالنسبة لبيانات الاختبار 57
2.3. نتائج المنهج رقم /2/ 58
1.2.3. دقة التصنيف بالنسبة لبيانات التدريب 58
2.2.3. دقة التصنيف بالنسبة لبيانات الاختبار 60
3.3. مقارنة نتائج المنهجين /1/ و/2/ 61
4.3. نتـــــائج انتقـــــــاء المعـــايير 64
4.3. نتـائج تجزئة بيانات التفقد اليومي 64

الفصل الرابع: النتائج والتوصيات 68
1.4. النتـــــائج 69
2.4. التوصيات 72
المراجـــــــــع 74
المـــــــــلاحق 77
الملخص باللغة الانكليزية 95

فهرس الجداول

الجدول 1 : أمثلة عامة عن إدارة الموارد البشرية المدمجة مع أدوات التنقيب في المعطيات ..... ق
الجدول 2: بعض تقنيات التنقيب في المعطيات الأصلية والهجينة المدمجة مع إدارة الموارد البشرية .. ل
الجدول 3 : الآليات المستخدمة في النموذج المقترح لإدارة المواهب البازغة ..... م
الجدول 4 : المعايير المعتمدة في النموذج المقترح لإدارة المواهب البازغة ..... ن
الجدول 5 : بعض المشاريع التطبيقية المدمجة مع تقنيات الذكاء الصنعي ..... ي
الجدول 6 : تقانات الذكاء الصنعي المستخدمة في ميدان إدارة الموارد البشرية ..... أأ
الجدول 1.2 : النموذج الموحد لتقييم أداء العاملين 41
الجدول 2.2 : نتائج خوارزميات التصنيف حسب المنهج رقم /1/ 48
الجدول 3.2 : النتائج المتوقعة لبيانات الاختبار حسب المنهج رقم /1/ 49
الجدول 4.2 : نتائج خوارزميات التصنيف حسب المنهج رقم /2/ 50
الجدول 4.2 : النتائج المتوقعة لبيانات الاختبار حسب المنهج رقم /2/ 51
الجدول 6.2 : نتائج خوارزميات انتقاء المعايير نتائج عملية تقييم أداء العاملين 51
الجدول 7.2 : توزع بيانات التفقد ضمن ثلاث مجموعات (k=3) 53
الجدول 8.2 : توزع بيانات التفقد ضمن أربع مجموعات (k=4) 53
الجدول 9.2 : توزع بيانات التفقد ضمن خمس مجموعات (k=5) 54
الجدول 1.3 : ترتيب نتائج خوارزميات التصنيف للمنهج /1/ حسب دقة التصنيف 55
الجدول 2.3 : النتائج المتوقعة لبيانات الاختبار حسب مبدأ تصويت الأكثرية 58
الجدول 3.3 : ترتيب نتائج خوارزميات التصنيف للمنهج /2/ حسب دقة التصنيف 59
الجدول 4.3 : النتائج المتوقعة لبيانات الاختبار حسب مبدأ تصويت الأكثرية 61
الجدول 5.3 : توزع بيانات المجموعة (B) مع تغير قيمة المتحول (K) 66

فهرس الأشـــكال

الشكل 1 : النموذج المقترح لإدارة علاقات الزبائن في بيئة مصرفية ..... ع
الشكل 2 : آلية عمل النموذج المقترح لإدارة المواهب البازغة ..... م
الشكل 3 : البنية الشجرية الناتجة عن عملية التصنيف لأنواع التغيب عن العمل ..... هـ
تقنيات التنقيب في المعطيات الأصلية والهجينة المدمجة مع إدارة الموارد البشرية 15
الشكل 2.1 : شجرة القرارات الناتجة عن التصنيف J48 17
الشكل3.1 : الطبقات الثلاث لخوارزمية MultiLayerPerceptron 19
الشكل 4.1 : بنية العصبون الصنعي الواحد 19
الشكل 5.1 : مثال على توزع القيم عند استخدام التصنيف KNN 21
الشكل 6.1 : أمثلة على توزع خطوط الفصل لتصنيف SVM الخطي 23
الشكل 7.1 : أمثلة على توزع خطوط الفصل لتصنيف SVM غير الخطي 23
الشكل 8.1 : هوامش المستوي الفاصل بين بيانات التدريب 24
الشكل 9.1 : خطوات عمل خوارزمية العنقدة القائمة على النقاط المركزية 28
الشكل 10.1 : الحالات المختلفة لعمل خوارزميات العنقدة القائمة على كثافة المعطيات 30
الشكل11.1 : صورة توضيحية لواجهة عمل تطبيق WEKA 33
الشكل 12.1 : مصفوفة الشك الخاصة بتطبيق WEKA 35
الشكل 1.2 : النموذج المقترح لمعالجة المعطيات التي تم جمعها /البيانات الأصلية/ 46
الشكل 2.2 : مصفوفات الشك الناتجة عن تطبيق المنهج رقم /1/ حسب خوارزميات التصنيف المعتمدة 48
الشكل 3.2 : مصفوفات الشك الناتجة عن تطبيق المنهج رقم /2/ حسب خوارزميات التصنيف المعتمدة 50
الشكل 1.3: مصفوفات الشك للمنهج رقم /1/, مرتبة ترتيباً تنازلياً 57
الشكل 2.3 : مصفوفات الشك لنتائج تصنيف المنهج /2/, مرتبة ترتيباً تنازلياً 60
الشكل 3.3 : مقارنة نتائج دقة تصنيف المنهج /1/ مع دقة تصنيف المنهج /2/ 62
الشكل 4.3: توزع البيانات وتشكل المجموعات مع تعديل قيمة عدد المجموعات (K) 65
الشكل 4.1 : النموذج المقترح للإدارة الذكية للموارد البشرية 71

فهرس الملاحق

الملحق 1 : هرم المعرفة 77

الملحق 2 : نماذج تقييم الأداء الخاصة بالشركة السورية للتبغ 78

الملحق 3 : نماذج تقييم الأداء الخاصة بالمؤسسة العامة لسد الفرات 80

الملحق 4 : نماذج التقييم (أ, ب, ج) المصدقة من قبل وزارة الشؤون الاجتماعية والعمل ... 82

الملحق 5 : مثال تطبيقي للاستفادة من الحزمة البرمجية الخاصة بتطبيق WEKA 85

الملحق 6 : بيانات التدريب بصيغة ملف ARFF حسب المنهج /1/ 87

الملحق 7 : بيانات الاختبار بصيغة ملف ARFF حسب المنهج /1/ 90

الملحق 8 : بيانات التدريب بصيغة ملف ARFF حسب المنهج /2/ 91

الملحق 9 : بيانات الاختبار بصيغة ملف ARFF حسب المنهج /2/ 92

الملحق (10) – بيانات التفقد اليومي المدمجة مع المعلومات الأساسية 93

الملخـــــص

- مقدمة

في ظل التركيب الديناميكي لأسواق أعمال المجتمعات الحديثة وما يرافق ذلك من تغييرات مستمرة على كافة الصعد الاقتصادية والسياسية والاجتماعية يعتبر الكادر البشري أحد أهم ركائز نجاح المؤسسات بما يمتلكه من مهارات وكفاءات تؤهله للوقوف في وجه التحديات والمخاطر المتصاعدة.

ومع التطور المذهل في تكنولوجيا المعلومات خلال النصف الثاني من القرن السابق فقد ظهرت العديد من التطبيقات والبنى التقنية التي تهدف للمساعدة في إدارة الكوادر البشرية, هذه التطبيقات التي ساهمت ولا تزال تساهم في الإدارة الفعالة لأغلب فعاليات ونشاطات الكوادر البشرية اعتمدت خوارزميات وطرائق متعددة ابتداءً من عمليات الأرشفة والتنظيم إلى العمليات الإحصائية والتحليلية وصولاً إلى خوارزميات الذكاء الصنعي وآليات استكشاف المعرفة والتنبؤ المستقبلي.

- الهدف من البحث

يرمي البحث بشكل أساسي نحو تحقيق الأهداف الثلاثة التالية:

- أولاً: دراسة وتحليل كل من فعاليتي تقييم أداء العاملين والتفقد اليومي في الجمهورية العربية السورية كنماذج لفعاليات الكوادر البشرية التي يمكن الاستفادة من بياناتها ضمن تطبيقات الذكاء الصنعي.
- ثانياً: دراسة وتحليل آليات عمل أحد آليات الذكاء الصنعي وهي التنقيب في المعطيات واختيار الأدوات المناسبة لمعالجة بيانات فعاليات الكوادر البشرية.
- ثالثاً: اقتراح نموذج تطبيقي للإدارة الذكية للموارد البشرية يستند إلى التنقيب في المعطيات لمعالجة عينة البيانات التي تم جمعها بغية استكشاف المعرفة الكامنة داخل هذه المعطيات بما يدعم عملية

اتخاذ القرارات المتعلقة بالكوادر البشرية في مؤسسات الجمهورية العربية السورية ومن ثم مقارنة أداء التقنيات التي تم استخدامها وبيان مدى ملاءمتها لهذا النوع من المعطيات.

- أسلوب البحث

انتهجت الدراسة أسلوب البحث التحليلي التطبيقي من خلال دراسة وتحليل بعض فعاليات الكوادر البشرية (عملية تقييم أداء العاملين, عملية التفقد اليومي) وأدوات التنقيب في المعطيات (التصنيف, اختيار المعايير, العنقدة) ومن ثم اقتراح النموذج الملائم لطبيعة العمل في الجمهورية العربية السورية بالاستناد إلى معطيات حقيقية تم جمعها من بعض المؤسسات والشركات ومناقشة النتائج التي تم الحصول عليها.

- نتائج البحث

- يشكل التنقيب في المعطيات مرتكزاً أساسياً في الإدارة الذكية للموارد البشرية.
- تشكل معطيات فعاليات الكوادر البشرية منبعاً هاماً للمعرفة مما يستوجب إعطاءها الأهمية المطلوبة.
- النموذج المقترح قادر على معالجة المعطيات واستكشاف المعرفة بالشكل المطلوب ويمكن اعتباره نواة قابلة للتوسع لتشمل فعاليات وأنشطة أخرى للكوادر البشرية.

- قيمة البحث

دراسات عديدة ومتنوعة حاولت الدمج بين إدارة الموارد البشرية بشتى نشاطاتها مع تقنيات الذكاء الصنعي لاختيار المناسب منها لهذا النوع من المعطيات. تمتاز هذه الدراسات بحداثتها ومحاولاتها الحثيثة لإيجاد معايير لهذا الدمج, هذا على المستوى العالمي أما على المستوى الوطني فالأبحاث التطبيقية لتقنيات التنقيب في المعطيات قليلة نسبياً ولاتزال في طور الدراسات النظرية ومن هنا تنبع أهمية هذا البحث من خلال اقتراح نموذج لإدارة الموارد البشرية بالاستفادة من أدوات التنقيب في المعطيات.

- الكلمات المفتاحية

- اســـتكشاف المعرفة (Knowledge Discovery)
- إدارة الموارد البشرية (Human Resource Management)
- التنقـــيب في المعطيـــــات (Data Mining)
- الذكـــــاء الصـنعي (Artificial Intelligence)
- نظم دعم القرار (Decision Support System)

مقدمة (Introduction)

تعد إدارة الموارد البشرية اليوم أحد أهم الوظائف الإدارية التي لا تقل أهمية عن باقي الوظائف الأخرى كالإنتاج والعقود والتسويق وذلك لما ظهر من أهمية العنصر البشري ومدى تأثيره على الكفاءة الإنتاجية للمنشآت التي تكافح بغية تحقيق أهدافها من مرابح مالية ورفع جودة المنتج والخدمات المرتبطة وصولاً إلى دخول عالم الإبداع والتميز.[1] ومن البديهي القول بأن جميع هذه الأهداف مرتبطة بوجود العدد الكافي من الكادر البشري المميز المجهز بالكفاءات والخبرات المناسبة والموزع بشكل صحيح ضمن قطاعات العمل المختلفة.

في السنوات الأخيرة اتسع مفهوم إدارة الموارد البشرية ليشمل فعاليات رئيسية متعددة يأتي على رأسها تحليل وتوصيف الوظائف المرتبطة بأنشطة الموارد البشرية، تخطيط الموارد البشرية، انتقاء واختيار الكفاءات, اكتشاف المواهب البازغة, جذب واستقطاب الطاقات البشرية، تحفيز الموارد البشرية، تنمية وتدريب الموارد البشرية وغيرها من الأنشطة التي لاتزال تتجدد باستمرار. هذه الفعاليات عبارة عن عمليات غاية في الديناميكية يصعب قياسها وتقييمها بعض الأحيان وبالتالي فإن التحدي الأكبر لنظم المعلومات الجديدة يكمن في زيادة القدرة على تحليل هذه الأنشطة واستخراج المعلومات المفيدة من البيانات المتراكمة للمساعدة في اتخاذ القرارات بأسرع زمن ممكن وبأقل تكلفة.

إن تطبيقات إدارة أنشطة الموارد البشرية المدمجة مع تقنيات الذكاء الصنعي يمكن أن توفر الحل الأمثل لمعالجة البيانات المرتبطة بفعاليات الكوادر البشرية ويمكن اعتبار التنقيب في المعطيات أحد أهم هذه التقنيات التي تعمل بطريقة آلية أو شبه آلية على استكشاف المعرفة من خلال بناء النماذج وتحليل البيانات واكتشاف الأنماط والقواعد الحيوية بين المعطيات المرتبطة بالكوادر البشرية بغية تحديد الميول والاتجاهات والتخطيط المستقبلي.

[1] المنتدى العربي لإدارة الموارد البشرية, http://www.hrdiscussion.com.

مشكلة البحث (Research Problem)

المشكلة الأساسية للبحث تكمن في أن نظم معلومات الموارد البشرية تضم البيانات والمعلومات المرتبطة بفعاليات الموارد البشرية كجداول وفهارس تمهيداً لاستخدامها عبر الطرق الاحصائية التقليدية دون إدراك لأهمية المعطيات المخزنة وما تحتويه من معرفة مضمنة يمكن استخدامها للتنبؤ المستقبلي بسلوك المعطيات. ومع اتساع وتعدد فعاليات وأنشطة إدارة الكوادر البشرية فإن استكشاف المعرفة من خلال التنقيب في معطيات فعاليات الكوادر البشرية كتقييم أداء العاملين والتفقد اليومي يمكن أن تشكل مثالاً أولياً لإدراك أهمية هذه المعطيات. وفي القطر العربي السوري – كغيره من البلدان – هناك قوانين وأنظمة ضابطة لآلية تطبيق هذه الفعاليات بالإضافة لمحاولات جاهدة لإقحام البنى المعلوماتية المتقدمة في قطاع إدارة الموارد البشرية والتي تصطدم مع المعوقات الأربعة التالية:

1. الضعف في القدرة على جمع بيانات الكوادر البشرية ضمن بنى معلوماتية مناسبة.
2. التقليل من أهمية العمليات الدورية لجمع معطيات مرتبطة بفعاليات الكوادر كعملية تقييم الأداء السنوية ونصف السنوية أو عملية التدوين اليومي لالتزام العاملين بساعات العمل.
3. عدم التطرق لأساليب المعالجة المتقدمة للبيانات والاقتصار على التقنيات التقليدية من أرشفة وتخزين.

وعليه يمكن طرح المشكلة بدقة أكبر من خلال الأسئلة الثلاث التالية:

1. هل يمكن الاستفادة من تقنيات التنقيب في المعطيات لبناء نموذج تنبؤي قادر على التكهن بالوظيفة المناسبة أو الأداء المستقبلي للكوادر البشرية بالاعتماد على النتائج السابقة لتقييم الأداء؟
2. هل يمكن الاستفادة من تقنيات التنقيب في المعطيات للقيام بتحليل متقدم لبيانات التفقد اليومي ومحاولة استكشاف الروابط والعلاقات بين أفراد كادر العمل في المؤسسة؟
3. ماهي محددات وأفق وركائز نموذج ناجح لإدارة الموارد البشرية قائم على التنقيب في المعطيات بالاعتماد على معطيات فعاليات الكوادر البشرية؟

الدراسات المرجعية (Literature Review)

- الدراسات المرجعية باللغة العربية

1) دراسة, (خلوف وآخرون – 2009)[2] بعنوان (استخدام آليات التنقيب في المعطيات للمساعدة في اكتشاف عمليات الاحتيال في البيئة المصرفية) في سياق رسالة الدكتوراه للطالب فادي خلوف وتهدف إلى:

1. توضيح مفهوم التنقيب في المعطيات وتطبيق أدواته للحصول على المعرفة المطلوبة. حيث وضح المقال أن ازدياد الحاجة إلى تحليل المعطيات واكتشاف النماذج الجديدة التي تقدمها ترافقه تطور تطبيقات التنقيب في المعطيات وتنوع التقنيات التي تستند إليها والأهداف التي تحققها.
2. استعراض آلية جديدة لاكتشاف الاحتيال في بيئة مصرفية من خلال مقارنة نتائج نسبة التحذيرات الخاطئة لكل من خوارزمية (Naive Bayes) وخوارزمية آلة شعاع الدعم الخاصة (SVM) لمساعدة المصارف في اتخاذ القرار لتحديد كون الزبون يقوم بصفقة طبيعية أم احتيالية.

- بغية تحقيق أهداف البحث قام الباحث باتباع الخطوات التالية:

1. دراسة تحليلية لآلية عمل أنظمة الدفع الالكتروني وتصنيف العمليات التي تتم بشكليها القانوني وغير القانوني أي الاحتيالي منها.
2. دراسة إسقاط ناتج عمليات أنظمة الدفع الالكتروني القانوني وغير القانوني على نماذج تصنيف التنقيب في المعطيات من خلال مصفوفة الشك (Confusion Matrix) .
3. اعتماد كل من المصنفين (Naive Bayesian, SVM) لتصنيف البيانات التدريبية وهي عبارة عن عينة من قاعدة بيانات المصرف التجاري السوري تشمل /400,000/ سجل لعميات دفع الكتروني باستخدام آلات الصراف الآلي (ATM) وتقسيم هذه المعطيات إلى معطيات تدريب واختبار.

[2] خلوف, رزوق, آشميس, استخدام آليات التنقيب في المعطيات للمساعدة في اكتشاف عمليات الاحتيال في البيئة المصرفية, مجلة جامعة دمشق للعلوم الهندسية المجلد الخامس والعشرون, العدد الثاني, 2009.

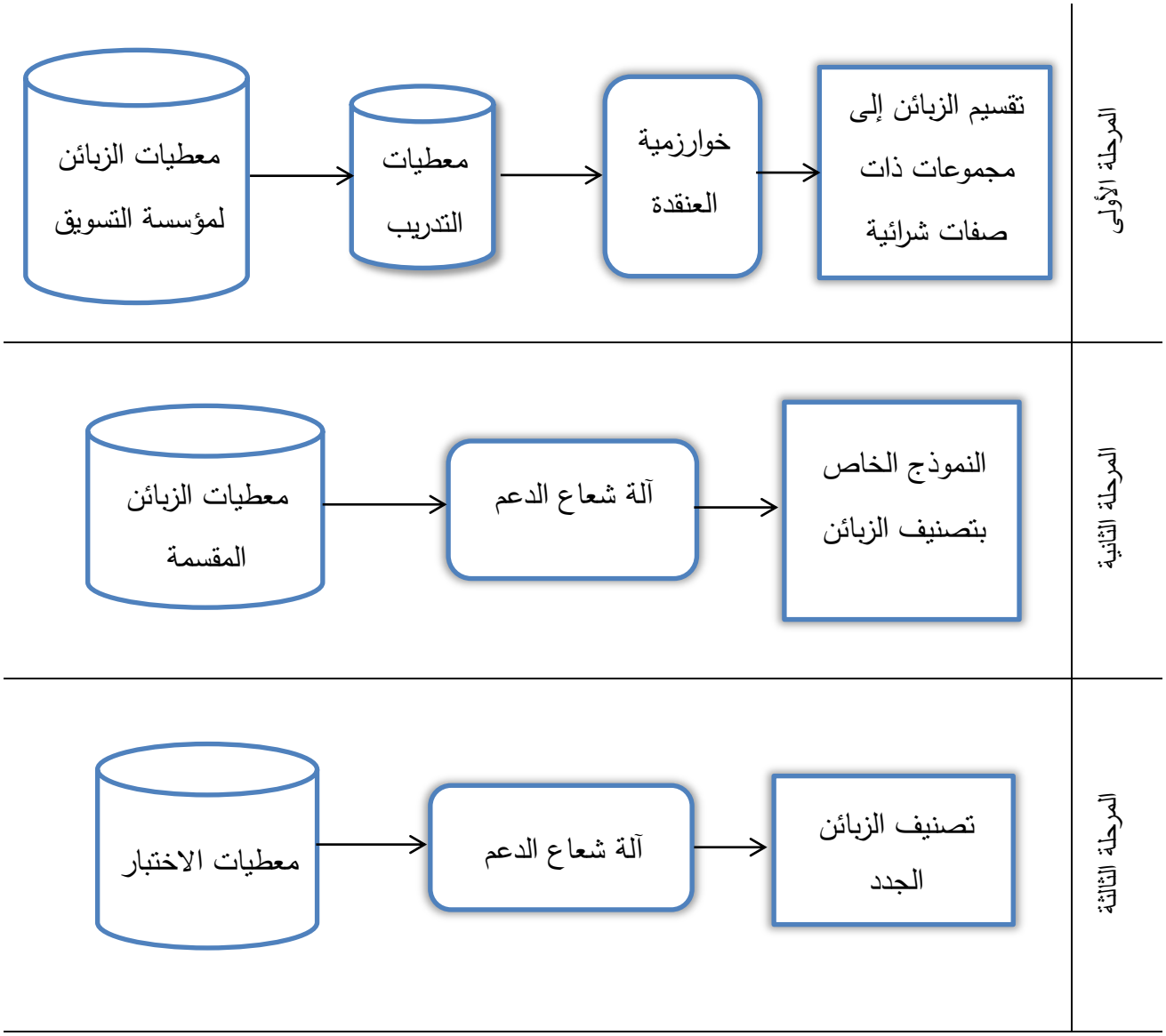
4. اتباع الخوارزمية التالية: معطيات التدريب ← تدريب المعطيات حسب الخوارزميتين السابقتين ← اختبار معطيات الاختبار ← مقارنة النتائج ← دمج نتائج التصنيفين ← الحصول على مصنف ذو دقة عالية قادر على كشف العمليات الاحتيالية وتمييزها عن العمليات النظامية.

- في النهاية يوصي البحث بتطبيق مصنفات أخرى لتحديد مدى تأثيرها وإضافة قواعد أخرى بغية اكتشاف الشذوذ في المعطيات وكقيام العميل بسحب لمبلغ من المال أكبر من مجموع الكميات المسحوبة خلال فترة زمنية محددة (أسبوع مثلاً).

2) دراسة, (خلوف وآخرون – 2010)[3] بعنوان (تطوير آليات جديدة للتنقيب في المعطيات لإدارة علاقات الزبائن في بيئة مصرفية). قدمت هذه المقالة توضيحاً عن مفهوم إدارة علاقات الزبائن وبشكل خاص التقطيع وتشكيل الزبائن وطريقة استخدام أدوات التنقيب في المعطيات لدعم عملية اتخاذ القرارات من خلال توضيح الخطوات المستخدمة لتوقع سلوك الزبائن بالاعتماد على تجميع المعطيات وإعدادها ومن ثم إيجاد الزبون المحتمل وأهمية هذا العمل في زيادة الأرباح.

- تم اقتراح النموذج المبين في الشكل (1) بغية انجاز العمل المطلوب بحيث يتم في المرحلة الاولى تقسيم الزبائن إلى مجموعات ذات صفات شرائية مشتركة وفي المرحلة الثانية يتم بناء نموذج خاص بتصنيف الزبائن من معطيات الزبائن التي تم تقسيمها في المرحلة الأولى وذلك باستخدام تصنيف آلة شعاع الدعم (SVM) وفي النهاية يتم في المرحلة الثالثة استخدام هذا النموذج لتصنيف الزبائن الجدد.

[3] خلوف, رزوق, شميس, تطوير آليات جديدة للتنقيب في المعطيات لإدارة علاقات الزبائن في بيئة مصرفية, مجلة جامعة دمشق للعلوم الهندسية المجلد السادس والعشرون, العدد الاول, 2010.



الشكل 1 : النموذج المقترح لإدارة علاقات الزبائن في بيئة مصرفية.

- النموذج المقترح يستخدم كلاً من خوارزميتي (K-Means) للعنقدة وآلة شعاع الدعم (SVM) للتصنيف, حيث أظهرت النتائج بأن دمج الخوارزميتين معاً أدى إلى تحسين دقة النموذج المقترح ورفعها إلى نسبة (81.35%) في تصنيف الزبائن الجدد الذين قد يزيدون من إنفاقهم في حال تم منحهم بطاقات الحسم مقابل (79.9%) عند استخدام آلة شعاع الدعم بشكل منفرد.
- تقترح الدراسة التعمق أكثر في هذا الموضوع والعمل على تعديل المتغيرات كقيمة العدد (k) المرتبط بخوارزمية (k-Means) أو استخدام خوارزميات تصنيف أخرى ومقارنة النتائج مع نتائج هذه الدراسة.

3) دراسة, (أحسن وعمار – 2009)[4] بعنوان (التنقيب في البيانات واتخاذ القرارات).

- أعدت الدراسة في سياق الملتقى الوطني السادس حول الأساليب الكمية ودورها في اتخاذ القرارات الإدارية 27-28 كانون الثاني 2009, جامعة سكيكدة, الجزائر.
- تقدم هذه الدراسة ملخصاً بسيطاً عن التنقيب في المعطيات وتتألف من عدة محاور هي:

1. مفهوم التنقيب في المعطيات: حيث أوردت الدراسة عدة تعاريف لهذا المفهوم منها "الاستكشاف الآلي أو المؤتمت لأنماط مخفية وغير جلية في قاعدة بيانات معينة".
2. العوامل المحركة لثورة التنقيب في المعطيات: حيث قسمت هذا العوامل إلى عوامل مرتبطة بجانب العرض كتأثيرات تطورات تكنولوجيا المعلومات وظهور أساليب تحليل جديدة من أبرزها الشبكات العصبونية, وعوامل مرتبطة بجانب الطلب كتنامي الحاجة الى نتائج تحليلية سريعة وانحسار التنظيم الهرمي حيث أصبح لزاماً على المدراء الاعتماد على أنفسهم في الحصول على احتياجاتهم من المعلومات التحليلية.
3. أدوات التنقيب في المعطيات كالتلخيص (Summarization) والتصنيف (Classification) والتنبؤ(Prediction) والتجزئة (Clustering) وغيرها من هذه الأدوات.
4. مراحل تطبيق عملية التنقيب في المعطيات التي أوجزها الباحثان بما يلي:

- فهم طبيعة الأعمال وكيفية تحقيق الفائدة القصوى من عمليات التنقيب واستكشاف المعرفة.
- فهم البيانات من خلال تجميع البيانات وتوصيفها وبيان جودتها.
- تهيئة البيانات لعمليات التنقيب من خلال صياغة المتغيرات والمعايير وضمان تكامل هذه البيانات وتناسقها وعدم وجود تعارض فيما بينها.
- صياغة نماذج الحل ذات الأداء المناسب.

[4] أحسن, طيار, التنقيب في البيانات واتخاذ القرارات, الملتقى الوطني السادس, جامعة سكيكدة, الجزائر, 27-28 كانون الثاني 2009.

- التقييم وتعليل نتائج النماذج.
- نشر وتوزيع النماذج داخل المنشأة الساعية لمساعدة عملية اتخاذ القرار.

5. مجالات تطبيق التنقيب في المعطيات: حيث سرد الكاتب المجالات الواسعة لاستخدام هذه التقنيات في شتى مجالات العمل.

4) رسالة جامعية, (الشربجي – 2006)[5], بعنوان (اكتشاف أنماط هامة في بيانات الطلاب في جامعة العلوم والتكنولوجيا بتطبيق تقنيات التنقيب عن البيانات).

- الباحث: أسماء عبد الغني قاسم الشرجبي, الدرجة العلمية: ماجستير, الجامعة: العلوم والتكنولوجيا, الكلية: العلوم والهندسة, القسم: الحاسوب, بلد الدراسة: اليمن, لغة الدراسة: الإنجليزية, تاريخ الإقرار: 2006.
- الهدف الرئيس لهذه الدراسة هو اكتشاف الأنماط (Patterns) الهامة والحيوية في بيانات طلاب جامعة العلوم والتكنولوجيا خلال 12 عاماً, ومن ثم استنباط مجموعة من السياسات العامة التي تكرس وتحسن الأداء الأكاديمي وتؤثر في اتخاذ القرار في الجامعة بشكل خاص، وفي مؤسسات التعليم العالي بشكل عام.
- تمحورت نتائج البحث والأنماط المكتشفة حول ثلاث قضايا رئيسة هي: تسرب الطلاب، فشل الطلاب، والتهرب من دفع الرسوم.
- أوضحت النتائج جملة من الأسباب المرتبطة بهذه القضايا والتي ترجمت فيما بعد لسياسات مقترحة لتحسين الأداء.

[5] أسماء عبد الغني قاسم الشربجي, 2006, اكتشاف أنماط هامة في بيانات الطلاب في جامعة العلوم والتكنولوجيا بتطبيق تقنيات التنقيب عن البيانات, أطروحة ماجستير غير منشورة, كلية العلوم والهندسة, جامعة العلوم والتكنولوجيا, اليمن, 2006.

- مثلت النتائج النهائية في هيئة قواعد قرار مع نسبة ثقة مرتبطة بكل قاعدة. ومن ثم قيمت النتائج في اتجاهين الأول بعرضها على متخذي القرار في إدارة الجامعة، وقد حققت نتائج جيدة في ضوء مجموعة من المعايير هي: الصحة ،الواقعية، الجدية والوضوح. أما الاتجاه الثاني فكان باختبار القواعد ومدى دقتها بشكل فعلي في مستودع البيانات باستخدام لغة الاستفسارات المهيكلة (SQL).

- الدراسات المرجعية باللغة الانكليزية

1) مقالة علمية, (Jantan et al – 2010) [6] بعنوان: (الاستفادة من خوارزميات تصنيف التنقيب في المعطيات للتنبؤ المستقبلي بأداء الموظف).

- تحاول هذه المقالة تسليط الضوء على استخدام خوارزميات التنقيب في المعطيات لاستخلاص النماذج من قواعد البيانات الموجودة بغية استخدامها في التنبؤ بالأداء المستقبلي لكوادر العمل.
- وضحت المقالة أن استخدام تقنيات التنقيب في المعطيات لإدارة الموارد البشرية لازال محدوداً وتم عرض بعض الأبحاث والدراسات السابقة في هذا المجال موضحة في الجدول (1).

الجدول 1 : أمثلة عامة عن إدارة الموارد البشرية المدمجة مع أدوات التنقيب في المعطيات.

أداة التنقيب في المعطيات	الاستخدام ضمن إدارة الموارد البشرية
التصنيف	انتقاء الكوادر البشرية (Personnel Selection, 2008) تحليل سلوك ووضع العمل (Job Attitudes, 2005) التوظيف واستثمار المواهب البازغة (Recruit & Retain Talents, 2007)
قواعد الارتباط	التدريب والتأهيل (Training, 2007)
التصنيف والتوقع	إسناد المهام للفعاليات المناسبة (Project Assignment, 2006)
التصنيف وقواعد الارتباط	انتقاء الكادر البشري (Personnel Selection, 2005)

[6] Jantan,, Hamdan, Othman, Puteh, 2010, Applying Data Mining Classification Techniques for Employee's Performance Prediction, kmice.cms.net.my, http://www.kmice.cms.net.my/ProcKMICe/KMICe2010/Paper/-PG601-607.pdf

- تم في هذه الدراسة استخدام تقنيات التصنيف التالية:
  - أشجار القرار (Decision Tree).
  - الشبكات العصبونية (Network Neural).
  - الجار الأقرب (Nearest Neighbor).
- عينة بيانات التدريب عبارة عن سجلات بيانات الكوادر البشرية التي تم جمعها على مدى ست سنوات وتتألف من 655 سجل.
- معايير التصنيف التي أخذها بعين الاعتبار لبناء نموذج قادر على التنبؤ بالأداء المستقبلي للعامل هي (التصنيف الوظيفي, الجنس, الشهادة, ناتج العمل, الكفاءات, المهارات الخاصة, الفعاليات, علامة التقييم, فئة التصنيف)
- تراوحت دقة التصنيفات بعد اجراءها على المعطيات بين 70% إلى 80%, حيث حازت تصنيفات أشجار القرار والجار الأقرب على الدقة الأعلى بين كانت نتائج الشبكات العصبونية ذات دقة أدنى.
- يوصي البحث في النهاية باستخدام تصنيفات أشجار القرار باعتبارها ذات أداء جيد لهذا النوع من لمعطيات كما يوصي باستخدام تصنيفات أخرى واختبار بيانات ومعطيات بأعداد أكبر لتبيان مدى كفاءة النتائج.

2) مقالة علمية, (Jantan et al - 2010)[7] بعنوان: (استخدام خوارزميات تصنيف التنقيب في المعطيات للتكهن بالمواهب البازغة)

- أكد البحث على أهمية استخدام تقنيات التنقيب في المعطيات لاستخراج المعلومات المرتبطة بإدارة المنابع البشرية من خلال استعراض بعض التقنيات المستخدمة في هذا الحقل وتطبيقاتها العملية كما

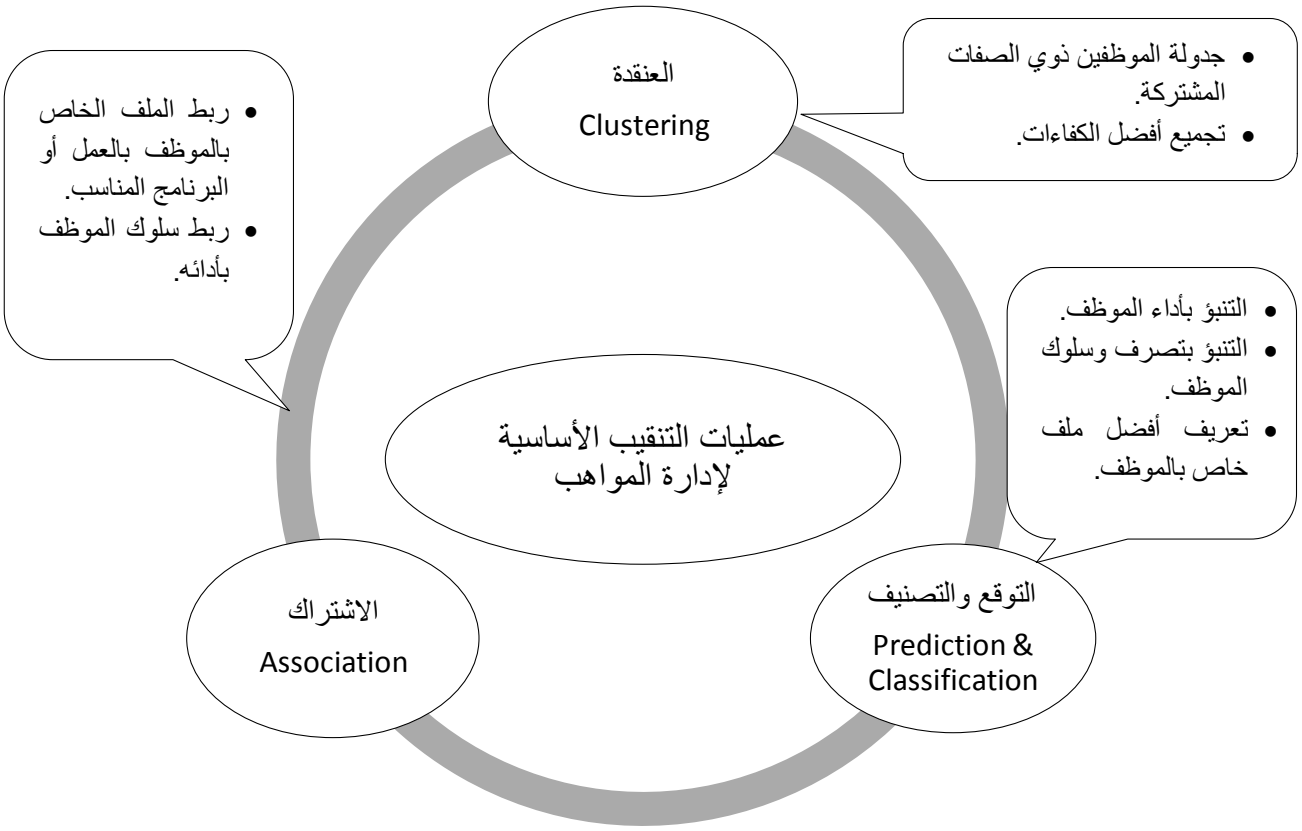
[7]13 .Jantan, Hamdan, Othman, 2011, Data Mining Classification Techniques for Human Talent Forecasting, Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1, InTech, DOI: 10.5772/14007. Available from: http://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining-/data-mining-classification-techniques-for-human-talent-forecasting

يبين الجدول (2) حيث يظهر أن عدد الأبحاث المطبقة في ميدان استخدام التقنيات المختلفة للذكاء الصنعي لاستكشاف المعرفة المتعلقة بموظفي المؤسسات أو تلك التي تساعد في دعم قرار المدراء قليل نسبياً.

الجدول 2 : بعض تقنيات التنقيب في المعطيات الأصلية والهجينة المدمجة مع إدارة الموارد البشرية.

الحقل التطبيقي	تقنيات التنقيب في المعطيات
اختيار الموظف (Personal Selection)	• أشجار القرار (Decision Tree). • التنقيب في المعطيات والمنطق العائم (Fuzzy Logic & Data Mining). • نظرية المجموعة التقريبية (Rough Set Theory).
التدريب (Training)	• التنقيب في قواعد الارتباط (Association Rule Mining)
تطوير قدرات الموظفين (Employee Development)	• التنقيب العائم في المعطيات (Fuzzy Data Mining) • الشبكات العصبونية المستندة للمنطق العائم (Fuzzy Artificial Neural Network). • أشجار القرار (Decision Tree).
تقييم الأداء (Performance Evaluation)	• أشجار القرار (Decision Tree).

- يركز البحث على تحديات إدارة المواهب البازغة في المؤسسات كالقدرة على اكتشاف تلك المواهب استناداً إلى تعريف دلائل ومعايير التي تميز بعض الكوادر عن غيرها.
- اقترح البحث النموذج الموضح في الشكل (2) المؤلف من حلقة تكرارية تبدأ من مرحلة تجميع الموظفين ذوي الصفات المشتركة فيما بينهم وتجميع الكفاءات ضمن مجموعات خاصة ومن ثم الانتقال إلى مرحلة التصنيف والتنبؤ للقيام بتصنيف البيانات التي تم تجميعها في المرحلة السابقة وبناء النماذج الخاصة بالصنفات المستخدمة وتجهيز هذه البيانات وإرسالها إلى المرحلة الثالثة التي تعمل على ربط الموظفين وخاصة المواهب البازغة بأماكن أو برامج العمل ذات المواصفات الخاصة والمتواجدة ضمن الموظف المناسب لهذه المهمة أو العمل.



الشكل 2 : آلية عمل النموذج المقترح لإدارة المواهب الباغة.

- تم اعتماد تقنيات التصنيف المبينة في الجدول (3) بغية إجراء عمليات التصنيف والتنبؤ.

الجدول 3 : الآليات المستخدمة في النموذج المقترح لإدارة المواهب البازغة.

تقنيات التنقيب في المعطيات	تقنية التصنيف
أشجار القرار	• C4.4 • Random Forest
الشبكات العصبونية	• MultiLayerPerceptron • Radial Basic Function Network
الجار الأقرب	• K*Star

- عينة بيانات التدريب مرتبطة ببيانات موظفي أحد الأكاديميات التدريسية وقد تم اعتماد المعايير الموضحة في الجدول (4).

الجدول 4 : المعايير المعتمدة في النموذج المقترح لإدارة المواهب البازغة.

المعايير	العدد	التوضيح
البيانات الأساسية	7	العمر, الجنس, ميدان العمر, عدد سنوات الخدمة, تاريخ الترقية الأولى, تاريخ الترقية الثانية ... الخ
التقييمات السابقة	15	درجات التقييم لآخر /15/ سنة سابقة.
المهارات والمعارف	20	المهارات الخاصة كالتعليم, الإشراف, الأبحاث, النشر, المؤتمرات ..
مهارات الإدارة	6	انضباط الطلاب والمهام الإدارية
ميزات متفرقة	5	المكافآت والعقوبات

- خلصت الدراسة إلى أن تصنيفات أشجار القرار بأنواعها ذات نتائج مقبولة ويمكن أن يحرز استخدامها في ذات الميدان مستقبلاً نتائج ممتازة إلا أن الدراسة أوصت باستخدام تقنيات أخرى واختبارها لا سيما تقنية (SVM) إضافة لاختيار معايير أخرى لإدارة الموارد البشرية.

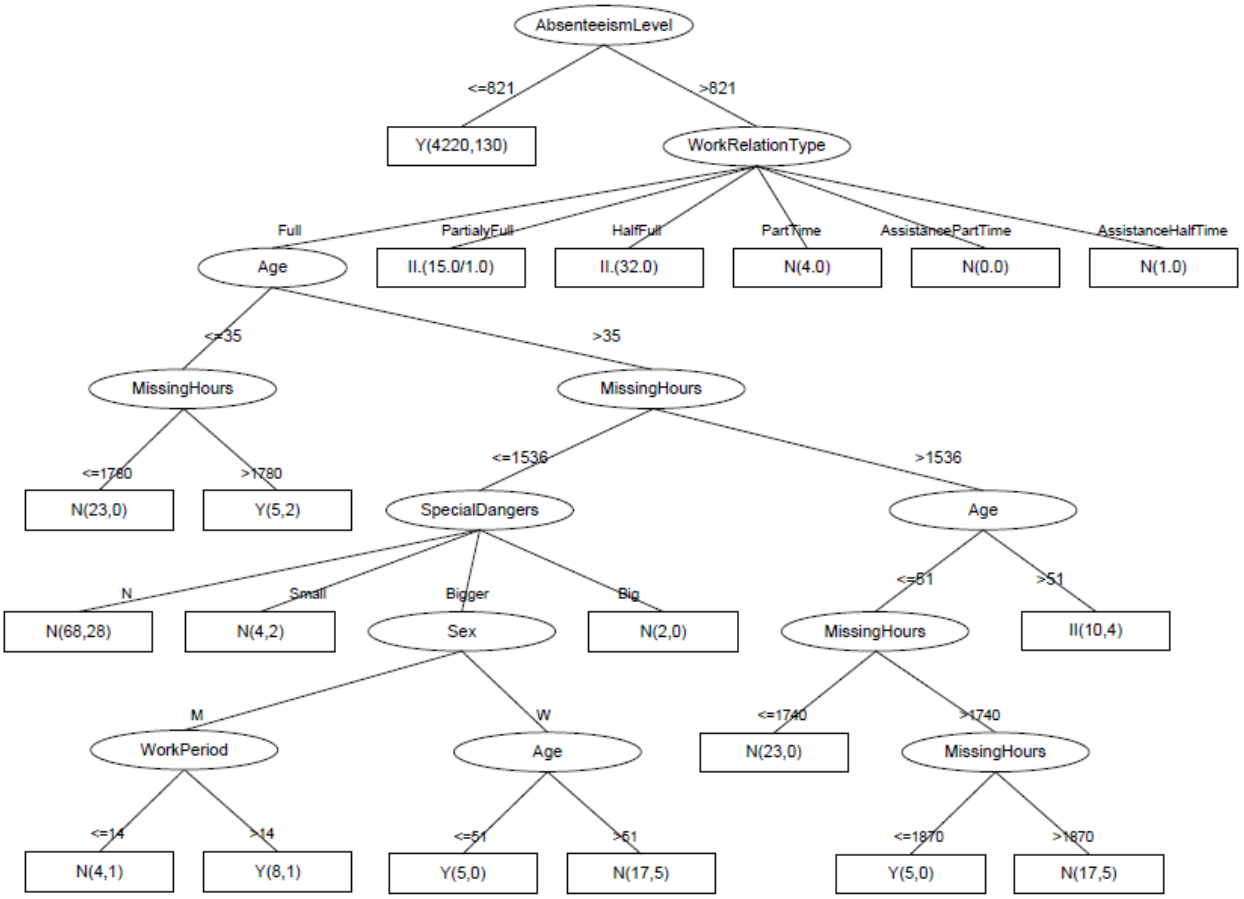
3) مقالة علمية, (Mojac & Igor – 2011)[8] بعنوان: (إدارة المعرفة وتكنولوجيا المعلومات في تحليل عمليات الموارد البشرية)

- تتطرق المقالة إلى موضوع الغياب غير المشروع للموظفين ومؤشرات هذا الغياب بالنسبة للوضع الوظيفي والوضع الاجتماعي للموظف. يعتمد العمل على تقسيم وتصنيف أنواع الغياب المتعارف عليها ضمن المؤسسات والمنظمات والتي تتراوح ما بين التأخر اليومي والإجازات الخاصة والإجازات الصحية والمرضية بالإضافة إلى الغياب غير المشروع.
- تكمن أهمية استخراج المعلومات المرتبطة بالتزام أو عدم التزام الموظف بأوقات العمل إلى فهم الصعوبات والمشاكل التي تعترض الموظف والتي يمكن أن تكون مشاكل متعلقة بشروط العمل كعدم الرضا عن العمل الحالي أو صعوبة شروط العمل أو المسافة البعيدة وشروط النقل السيئة والصعبة بين

[8] Mojca B., Igor B., (2011), Knowledge Management and Information Technology in Analyzing Human Resource Processes, The SPRING 5th International Conference on Knowledge Generation, Communication and Management: KGCM 2011, Available at http://www.iiis.org/CDs2011/CD2011IMC/KGCM_2011/-PapersPdf/GB772CV.pdf

مكان العمل ومكان السكن وغيرها من الصعوبات المرتبطة بالعمل. أو من الممكن أن المشكلات غير مرتبطة بالعمل وإنما ذات ارتباط بالوضع الاجتماعي أو العائلي للموظف ولا سيما المشكلات العائلية أو الالتزامات العائلية المجهدة.

- عمد الباحثون في هذا العمل إلى جدولة أنواع التغيب عن العمل والاستفادة من خوارزمية طريقة أشجار القرار في تصنيف المعطيات كما يوضح الشكل (3) والذي يبين أنواع التغيب وارتباطها مع المعايير.



الشكل 3 : البنية الشجرية الناتجة عن عملية التصنيف لأنواع التغيب عن العمل.

- خلصت الدراسة للتأكيد على إيلاء موضوع معالجة التغيب والتسرب عن أوقات الدوام الرسمي أهمية أكبر و الاستفادة من تقنيات الذكاء الصنعي المختلفة في تحليل هذا النوع من البيانات.

4) مقالة علمية, (Jantan et al -2010)[9] بعنوان: (تقنيات ذكية لدعم اتخاذ القرار النظام في إدارة الموارد البشرية)

- تقدم هذه الدراسة فكرة لتطبيق نهج الأنظمة الذكية في اتخاذ القرار (IDSS) ضمن حقل إدارة الموارد البشرية من خلال استخدام وتطبيق تقنيات الذكاء الصنعي الهجينة من خلال دمج عدد من هذه التقنيات مما يسمح لهذه الأنظمة باتباع السلوكيات التالية:

1. الإدراك والاستنتاج من الخبرات المتراكمة.
2. اتخاذ القرار في الحالات العائمة والغير مؤكدة.
3. القدرة على الاستدلال المنطقي.
4. القدرة على الاستدلال بشكل مبسط ورشيد.
5. الاستجابة السريعة والصحيحة للمواقف المتعاقبة.
6. تشخيص الأهمية المرتبطة بعناصر الحالات المختلفة.
7. إدراك وفهم الخطابات المتناقضة أو المبهمة.

- يوضح البحث أنواع نظم دعم القرار المتعارف عليها ويركز على الفئة الثانية من الفئات الثلاث:

1. نظم دعم القرار غير الفعالة (Passive DSS): وهي أنظمة تقليدية تستجيب للمتغيرات بناء على الاجراءات والقواعد التي تم تلقينها للنظام وبالتالي فإن مكونات هذا النوع من النظم عبارة عن قواعد بيانات وقواعد ارتباط ونظم تحليل ومعالجة.
2. نظم دعم القرار الفعالة (Active DSS): على عكس النظم السلبية فإن هذه النظم تتصف بقدرتها على التعلم واستنباط المعرفة كالنظم الخبيرة (Expert System) والنظم القائمة على

[9] Jantan, Hamdan, Othman, 2010, Intelligent Techniques for Decision Support System in Human Resource Management, Decision Support Systems Advances in, Ger Devlin (Ed.), ISBN: 978-953-307-069-8, InTech, DOI: 10.5772/39401. Available from: http://www.intechopen.com/books/decision-support-systems-advances-in/intellige-nt-techniques-for-decision-support-system-in-human-resource-management

المعرفة (Knowledge-Based System). هذا وتعتبر أدوات التعلم الآلي (Machine Learning) أهم مكونات هذا النوع من النظم.

3. نظم دعم القرار التفاعلية (Proactive DSS): والتي تعرف أيضاً باسم بالنظم المستندة إلى المحاسبات كلية الوجود (Ubiquitous Computing Technology-Based) المتضمنة نظم دعم القرار بالإضافة إلى عوامل تشخيص وإدراك المحتوى (Context Aware Functionalities). يتسم هذا النوع من النظم بسهولة التنقل (Mobility) والحمل (Portability) والحيوية (Activeness).

- تذكر الدراسة بعض التجارب التي تم خلالها دمج عدد من تقنيات الذكاء الصنعي بغية الحصول على نتائج مميزة, كما يبين الجدول (5) فإن العديد من المشاريع الطموحة تم تنفيذها من خلال دمج عدة تقنيات للذكاء الصنعي ولا سيما التنقيب في المعطيات والشبكات العصبونية والمنطق العائم

الجدول 5 : بعض المشاريع التطبيقية المدمجة مع تقنيات الذكاء الصنعي.

العمل المنفذ	تقنيات الذكاء الصنعي المدمجة
التنبؤ بحالة الطقس	Data mining & ANN
التوقع والتنبؤ بالمبيعات	ANN & Temporal Difference Learning method
قياس ودراسة واقع ميزان التسويق	Genetic Algorithm based Fuzzy Neural Network
التحكم بتلوث الهواء في مصانع الفحم الحجري	Fuzzy set and Gaussian dispersion model
الخيارات الاستراتيجية لمدى التداخلات التقنية في البنى التحتية الوطنية.	Case-based, mobile agent and multi-agent
تقدير مدى أداء المشاريع	Model based and Rule based
تقدير مدى التقدم المدني	Knowledge-based system and ANN
إدارة استيراد تقنيات المعلومات المؤثرة	Knowledge-based system and Fuzzy theory

- الدراسة الإطلاعية لهذا البحث قامت بتصنيف أبحاث استخدام تقنيات الذكاء الصنعي في إدارة المنابع البشرية ضمن أربعة فئات رئيسية مبينة في الجدول (6) ضمت انتقاء الموظفين الجدد أثناء عمليات التوظيف واختيار المكان المناسب لإمكانيات الموظف إضافة إلى جدولة تأهيل وتدريب الكوادر البشرية

واختيار الدورات التأهيلية المناسبة لرفع أداء العاملين, كما أن استخدام تقنيات الذكاء الصنعي تواجدت في عمليات التحفيز والإدارة.

الجدول 6 : تقانات الذكاء الصنعي المستخدمة في ميدان إدارة الموارد البشرية

الفئة	التقانات المستخدمة
التوظيف وانتقاء الموظف	Expert System/Knowledge-Based System, Data mining, Artificial Neural Network
التأهيل والتدريب	Knowledge-based system, Expert System, Rough Set Theory
التحفيز	Artificial Neural Network, Fuzzy logic
الإدارة	Software agent

- قامت الدراسة بالنهاية باقتراح نموذج للإدارة الذكية للمنابع البشرية يتألف من أربعة مكونات رئيسية هي:

1. تقنيات استكشاف المعرفة في قواعد البيانات (KDD): وذلك بهدف ايجاد نموذج تنبؤي واكتشاف الروابط الكامنة بين البيانات الموجودة ضمن قاعدة البيانات.
2. نظام إدارة النماذج (Model Management System): الذي يقوم بتخزين النماذج المنشأة والنظم المحاكاة والنماذج المترابطة.
3. نظام قواعد المعرفة (Knowledge Base System): والذي يتضمن مجموعة من الروابط والمعارف التي تعتمد عليها قواعد الربط.
4. نظام استشاري (Advisory System): عبارة عن محرك استنباطي للإشراف على تفاعل المكونات الثلاثة سابقة الذكر.

- ملخص الدراسات المرجعية (Literature review summery)

تباينت الدراسات المرجعية بين دراسات استقصائية ودراسات بحثية ودراسات تطبيقية اشتركت جميعها بعدد من النقاط أساسية:

1. تشكل أدوات التنقيب في المعطيات أحد أهم تقنيات الذكاء الصنعي التي يمكن الاستفادة منها في عمليات استكشاف المعرفة.
2. إن استخدام الأدوات المختلفة للتنقيب في المعطيات مرتبط بنمط العمليات المراد تنفيذها وأهدافها, وبالتالي يجب اختيار الأدوات المناسبة للمعطيات المراد تحليلها.
3. لم يحظى موضوع الاستفادة من التنقيب في المعطيات في مجال إدارة الكوادر البشرية بالاهتمام الكافي.
4. لا يوجد اتجاهات ومحددات خاصة للتنقيب في معطيات فعاليات الكوادر البشرية وجميع الآليات المقترحة مرتبطة بعينة المعطيات الخاصة بالمؤسسة التي تم اجراء البحث عليها.

- المقارنة بين الدراسة الحالية والدراسات المرجعية

(This Study VS Literature review)

تتشابه الدراسة الحالية مع الدراسات السابقة في تناولها لموضوع الاستفادة من تقنية التنقيب في المعطيات في إدارة الموارد البشرية ولكن الفرق الرئيسي يكمن في أن النموذج المقترح في هذه الدراسة معد للتطبيق بصورة عملية وليس لمجرد اختبار النتائج ويتجلى ذلك في نقطتين أساسيتين:

1. التعلم الذاتي للنموذج المقترح من خلال إضافة بيانات الاختبار التي حظيت بتوافق كامل في التنبؤ حسب مبدأ التصويت للأكثرية (Majority Voting Scheme) إلى بيانات التدريب وبالتالي فإن بيانات التدريب تنمو بشكل مستمر وبالتالي ارتفاع الدقة التنبؤية للنماذج.

2. اقتراح نموذج موحد لجمع نتائج تقييم أداء العاملين ومقارنة أداء خوارزميات التصنيف لاختيار المناسب منها لهذا النوع من المعطيات.

هذا بالإضافة إلى أن هذا البحث يمتاز بنقطة جوهرية تتجسد بكونه يلائم واقع القطر العربي السوري نظراً لأن جميع الفعاليات التي تمت دراستها إضافة للعينة التجريبية التي تم جمعها تعود لمؤسسات الجمهورية العربية السورية.

أهمية البحث (Research Importance)

1) المستوى العالمي: يشكل هذا العمل رديفاً لدراسات عالمية متعددة تحاول التأكيد على مدى أهمية التنقيب في المعطيات في الإدارة الذكية للموارد البشرية.
2) المستوى العربي: ينضم هذا البحث إلى عدة أبحاث عربية تسلط الضوء على التنقيب في المعطيات كأداة لاستكشاف المعرفة, علماً أنها التطبيق العملي الأول من نوعه على مستوى النشر العربي الذي يعمل على الدمج بين إدارة الموارد البشرية وتقنيات الذكاء الصنعي.
3) المستوى الوطني: تتجلى قيمة البحث وطنياً في محورين:

- المحور الأول يتضمن إظهار مدى أهمية كل من عمليتي تقييم الأداء والتفقد اليومي لما تتضمنه من معطيات قيّمة يمكن الاستفادة منها في التخطط المستقبلي للكوادر وتسهيل عملية اتخاذ القرار. كما تم في هذا البحث تحليل النماذج المستخدمة حالياً لإنجاز هاتين الفعاليتين واقتراح النماذج المناسبة لجمع نتائج ومعطيات هاتين الفعاليتين بحيث تكون جاهزة للاستخدام ضمن بيئة التنقيب في المعطيات.
- المحور الثاني من خلال توضيح مفهوم التنقيب في المعطيات والتقنيات المستخدمة للقيام بذلك بالإضافة لاقتراح نموذج تطبيقي للاستفادة من هذه التقنيات في إدارة الموارد البشرية وتنفيذه بشكل عملي بالاستفادة من بيانات حقيقية تم جمعها من بعض مؤسسات القطر العربي السوري.

أهدف البحث (Research Aim)

يرمي البحث بشكل أساسي لتحقيق الأهداف الثلاثة التالية:

- أولاً: دراسة وتحليل أساليب جمع معطيات كل من فعاليتي تقييم أداء العاملين والتفقد اليومي في بعض مؤسسات الجمهورية العربية السورية وتعديلها بحيث يمكن الاستفادة منها ضمن بيئة الذكاء الصنعي.

- ثانياً: دراسة وتحليل الأدوات المتعددة التنقيب في المعطيات ومن ثم اختيار الأدوات المناسبة لمعالجة عينة البيانات التي تم جمعها.
- ثالثاً: اقتراح نموذج لاستكشاف المعرفة الكامنة في معطيات كل من فعاليتي تقييم أداء العاملين والتفقد اليومي ومن ثم التنفيذ العملي لهذا النموذج بالاستفادة من عينة المعطيات التي تم جمعها من بعض مؤسسات الجمهورية العربية السورية ومقارنة تقنيات التنقيب التي تم استخدامها لاختيار المناسب منها لهذا النوع من المعطيات.

فرضيات البحث (Research Hypothesis)

1. ضعف استخدام نظم المعلومات ضمن حقل إدارة الكوادر البشرية وعدم القدرة على الوصول إلى معطيات خاصة بمؤسسة بحد ذاتها لاعتبار هذه المعطيات بيانات حساسة لا يمكن الكشف عنها.
2. العينة المراد إجراء البحث عليها هي عينة غير مرتبطة بمنشأة بحد ذاتها وإنما هي عينة عشوائية وحقيقية لمجموعة من العاملين في القطاع العام والخاص في الجمهورية العربية السورية تم جمعها وتنقيحها بالتعاون مع بعض الزملاء والمهتمين بالموضوع.
3. عدم التطرق إلى موضوع نزاهة تقييم الأداء أو مدى عدالته أو مشكلاته الإدارية, وبالتالي فإن التقييم المقصود هو تقييم موضوعي.

منهج البحث (Research Methodology)

يعتمد البحث المنهج التحليلي التطبيقي, ويتجلى ذلك من خلال المحورين التاليين:

- المنهج التحليلي: ويتضمن الاطلاع على أساليب ونماذج عملية تقييم أداء العاملين وعملية التفقد اليومي في سورية ودراستها وتحليلها بغية إيجاد الإمكانية للاستفادة من معطياتها في تطبيقات الذكاء الصنعي والتنقيب في المعطيات.

- المنهج التطبيقي: من خلال التطبيق العملي للنموذج الذي تم اقتراحه في هذا البحث ومن ثم مقارنة النتائج لتحديد أفضل هذه التقانات في هذا الميدان.

مجتمع البحث وعينته (Research Society and Sample)

- مجتمع البحث: فعاليات وأنشطة الكوادر البشرية في مؤسسات الجمهورية العربية السورية.
- عينة البحث: البيانات المرتبطة بكل من فعاليتي تقييم أداء العاملين والتفقد اليومي لعدد من العاملين.
- أداة البحث: تطبيق Weka, وهو تطبيق خاص بخوارزميات التنقيب في المعطيات والتعلم الآلي.

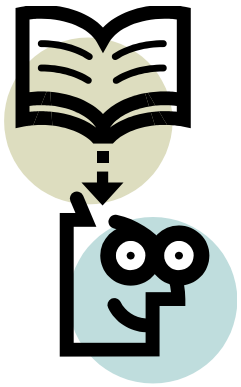
طريقة تنفيذ البحث (The Manner of Implement The Research)

- القسم النظري: يتضمن التعريف بفعاليتي تقييم أداء العاملين التفقد اليومي ومطالعة الأسس والمعايير الناظمة لتنفيذ هاتين الفعاليتين في الجمهورية العربية السورية وآلية جمع معطياتهما لدى بعض المؤسسات ومن ثم توضيح تمهيدي لمفاهيم ومبادئ التنقيب في المعطيات والأدوات التي سيتم استخدامها وأنواع الخوارزميات التي تستند إليها أثناء عملها.
- التطبيق العملي: يتضمن سرد الخطوات العملية المتعلقة بجمع المعطيات وتصميم نماذج مقترحة لجدولة عينة البيانات بالشكل الذي يسمح باستخدامها ضمن بيئة عمل Weka, من ثم اجراء خوارزميات التنقيب المتعددة لبناء النموذج المقترح لمعالجة معطيات فعاليات الكوادر البشرية وجدولة النتائج, وفي النهاية مناقشة النتائج لاستخلاص المقترحات والتوصيات.

الفصـــــــل الأول

التمهيد النظري

Theoretical Entrance



تمهيد:

يتناول هذا الفصل المفاهيم النظرية للبحث وفق أربعة محاور:

1.1. عملية تقييم أداء العاملين.

2.1. عملية التفقـــــد اليومـــي.

3.1. التنقيب في المعطيـــــات.

4.1. تطـــــــــــبيق Weka.

1.1. تقييم أداء العاملين (Employee's Performance Evaluation)

تعتبر عملية تقييم الأداء من الأدوات الهامة التي تستخدمها الإدارة في الرقابة على الموارد البشرية للمؤسسة, حيث حظي هذا الموضوع على اهتمام بالغ وشهد بحوثاً مستمرة لتحسين هذه العملية. يمكن تعريف عملية تقييم الأداء بالعملية التي تمكن المؤسسة من قياس كفاءة موظفيها وتحليل وتقييم أنماط ومستويات أدائهم وتعاملهم وتحديد درجة كفاءتهم الحالية والمتوقعة وبمعنى آخر الحصول على التغذية الرجعية لفعالية العاملين.[10]

1.1.1. نبذة عن تقييم أداء العاملين في سورية
(About Employee's Performance evaluation In Syria)

- يستند تقييم أداء العاملين في الجمهورية العربية السورية إلى المرسوم /322/ لعام 2005 المتضمن نظام تقييم أداء العاملين الخاضعين لأحكامه وذلك بناء على أحكام المادة 23 من القانون الأساسي للعاملين في الدولة[11].
- إنّ المقصود بتقييم الأداء للعاملين هو تحديد مستوى أداء للعامل في جميع المستويات الإدارية ولجميع الوحدات التنظيمية بالمقارنة مع معايير مرجعية خلال فترة زمنية محددة. [12]
- تتوجب الإشارة لاستخدام مصطلح تقييم أداء العامل وليس تقييم العامل. وبذلك لا يبحث هذا التقييم في الأمور الشخصية أو الخصائص الجسدية للعامل بل في إنتاجية العامل وانسجام مهاراته وكفاءاته مع البيئة المؤسسية التي يعمل بها، واستخدام هذه الكفاءات والمهارات في مكانها وزمانها الصحيحين. [13]
- إنّ تقييم أداء العامل كفرد يختلف عن تقييم أداء المؤسسة أو تقييم أداء منظومة ما في هذه المؤسسة (نظام القرارات، أو النظام المالي، أو نظام العقود ...الخ), ولكن عملية تقييم الأداء هي عملية يجب أن

---

[10] المنتدي العربي لإدارة الموارد البشرية.

[11] المرسوم التشريعي /322/ لعام 2005, تاريخ 2005/7/27,

[12] اليونس, 2007, تقييم أداء العاملين والترفيعات السنوية, ورشة عمل أقامتها مديرية التدريب والتأهيل والمعلوماتية في المؤسسة العامة لسد الفرات, سورية, ص 4

[13] المرجع ذاته.

تتكامل مع النظم والعمليات الأخرى للمؤسسة[14] (نظام المعلومات، نظام الاتصالات المادية والبشرية، النظام المالي والإداري وغيرها).

2.1.1. أهداف التقييم (Evaluation's Objectives)

حسب (المرسوم 322 – 2005)[15] فإن عملية التقييم تهدف إلى:

1. استخدام نتائج تقييم الأداء أساساً في اتخاذ قرارات الترفيع والنقل والتسريح والمكافآت وإنهاء التعاقد.
2. تخطيط الموارد البشرية وتحديد هيكل العمالة في المؤسسة.
3. تقويم سياسات وأساليب الاختيار والتعيين المستخدمة.
4. تحديد الاحتياجات التدريبية للعاملين.
5. تحديد مدى نجاح الأفراد في أداء أعمالهم وترشيد قرارات الإدارة المتعلقة بتحسين مستوى هذا الأداء في المستقبل.
6. وضع كل عامل في العمل الذي يتناسب مع قدراته ومؤهلاته وأدائه.
7. تقويم أسلوب القيادة والإشراف المستخدم.

3.1.1. نوع التقييم المستخدم (Evaluation Type)

تم اعتماد عملية تقييم الأداء نصف السنوية والتي تجرى كل ستة أشهر من قبل الرئيس المباشر والمدير الذي يعلوه مباشرة في المستوى الإداري كل على حدة[16]، وفق الإجراءات والقواعد المبينة لاحقاً.

---

[14] المرجع ذاته.

[15] المرسوم التشريعي /322/ لعام 2005, تاريخ 2005/7/27, الباب الثاني, المادة 2.

[16] المرسوم التشريعي /322/ لعام 2005, تاريخ 2005/7/27, الباب الثالث, المادة 3.

3.1.1.أ. تفصيلات التقييم نصف السنوي (Semi Annual Evaluation Details)

قامت اللجان المختصة في وزارة الشؤون الاجتماعية والعمل بتصميم نماذج تقييم الأداء وتوزيع فئات العاملين ضمن نماذج تقييم الأداء وفقاً لمعيار العمل الفعلي الذي يقوم به العامل عند التقييم وفق ثلاث وظائف رئيسة هي:

1. الوظائف الإشرافية والقيادية[17]:

يندرج تحت هذا الوصف جميع عاملي المؤسسات من كافة الفئات والمكلفين برئاسة ورشة أو شعبة فما فوق على أن يراعى في التقييم عدد وفئات العاملين محل القيادة والإشراف.

2. الوظائف الفنية والخدمية[18]:

أ- الوظائف الفنية التي تضم كلاً من:

1. جميع عاملي الفئة الأولى من الجامعيين غير المشمولين بالوظائف الإشرافية والقيادية.
2. جميع عاملي الفئة الثانية الفنيين من خريجي المعاهد الفنية أيا كانت هذه المعاهد من غير المشمولين بالوظائف الإشرافية والقيادية.
3. حملة الثانوية بكافة فروعها والمعينين على هذا الأساس وغير المشمولين بالوظائف الإشرافية والقيادية.
4. حملة الإعدادية بكافة فروعها والمعينين على هذا الأساس وغير المشمولين بالوظائف الإشرافية والقيادية

ب- الوظائف الخدمية التي تضم كلاً من:

1. عاملي الفئة الرابعة الذين لا يعملون في حقل الإنتاج المباشر إنما بالوظائف الخدمية مثل السائقين والمراسلين وغيرهم.

---

[17] وزارة الشؤون الاجتماعية والعمل, لجنة تقييم أداء العاملين, نماذج تقييم الأداء الخاصة بالوظائف الإشرافية والقيادية.

[18] المصدر ذاته, نماذج تقييم الأداء الخاصة بالوظائف الفنية والخدمية.

2. عاملي الفئة الخامسة من حراس ومستخدمين وعمال عاديين بدون جهد عضلي.

3. الوظائف الإنتاجية[19]: تنحصر الوظائف الإنتاجية في العاملين ضمن قطاع الإنتاج المباشر وعمال الجهد العضلي والمستفيدين من تعويض الصيانة والاعتناء ويقومون بأعمال الإنتاج الفعلية وهم على سبيل المثال لا الحصر عمال الانتاج لكافة منتوجات المؤسسات ذات الطابع الإنتاجي أو الاقتصادي إضافة إلى العمال الميكانيكيين والكهربائيين الذي يقومون بإصلاح الآليات والمعدات الإنتاجية ويقفون خلفها إضافة إلى عاملي الورش الإنتاجية الذين يولد جهدهم منتج وغيرهم.

3.1.1.ب. نقاط التقييم[20] (Evaluation Points)

- بغية حساب درجات ونقاط التقييم فقد تم تقسيم نماذج التقييم المعتمدة في القطر إلى البنود التالية:
  1. الكفاءات الأساسية.
  2. مهام الوظيفة.
  3. أهداف الوظيفة.
- إن مقياس نقاط تقييم كل عامل من العوامل المشار إليها أعلاه يتراوح بين (0 إلى 5) درجات ويتم الحساب وفق ما يلي:
  1. مجموع نقاط عوامل الكفاءات الأساسية مقسماً على عدد العوامل.
  2. مجموع نقاط عوامل مهام الوظيفة مقسماً على عدد العوامل.
  3. مجموع نقاط عوامل أهداف الوظيفة مقسماً على عدد العوامل.
  4. تجمع النتائج الثلاث السابقة وتقسم على /3/.
  5. تقرب الكسور عند احتساب معدل التقديرات في تقييم الأداء إلى اقرب رقميين عشريين.

---

[19] المصدر ذاته, نماذج تقييم الأداء الخاصة بالوظائف الخدمية.

[20] المرسوم التشريعي /322/ لعام 2005, تاريخ 2005/7/27, الباب الخامس, المادة 12.

- يتم تقدير الدرجات المحسوبة في الخطوة السابقة كما يلي :

1. من 1 – 2.49 نقطة, تساوي درجة ضعيف
2. من 2.50 – 3.49 نقطة, تساوي درجة وسط
3. من 3.50 – 5 نقطة, تساوي درجة جيد

4.1.1. نماذج تقييم الأداء المعتمدة (Approved Evaluation Forms)

إن جميع ما ذكر أعلاه موحد بين جميع المؤسسات الحكومية في الجمهورية العربية السورية ولكن الاختلاف القائم هو في معايير عملية التقييم كما ذكر في المرسوم /322/:

"تقوم كل جهة عامة من خلال لجنة مختصة تشكل بقرار من الوزير المختص أو المحافظ أو المدير العام أو من في حكمهم, بوضع نماذج خاصة بتقرير تقويم الأداء لكل نوع من الوظائف حسب طبيعة نشاطها وطبيعة مهام الوظيفة, وذلك وفق الإرشادات والنماذج الإرشادية المرفقة بهذا المرسوم, على أن يتم اعتماد هذه النماذج من قبل لجنة مختصة تشكل لهذا الغرض بقرار من وزير الشؤون الاجتماعية والعمل قبل البدء باستخدامها من قبل الجهة العامة"[21]

وبالتالي تعمد المؤسسات إلى وضع المعايير المناسبة لطبيعة عملها وطبيعة عمل موظفيها ومن ثم تقوم بتصديقها من وزارة الشؤون الاجتماعية والعمل – لجنة تقييم أداء العاملين, وبعد عملية التصديق تصبح نماذج معتمدة, وفيما يلي أمثلة عن هذه النماذج مأخوذة من عدة مؤسسات وشركات حكومية:

أ. الشركة العامة للتبغ

- كما هو مبين في الملحق رقم (1) فإن معايير التقييم الخاصة بالوظائف الإنتاجية والخدمية مؤلفة من /19/ معياراً بينما مقسمة كما يلي:
  - الكفاءات الأساسية, وتضم /11/ معياراً.
  - مهام الوظيفة, وتضم /3/ معياراً.

[21] المرسوم التشريعي /322/ لعام 2005, تاريخ 2005/7/27, الباب الخامس, المادة 11.

- أهداف الوظيفة وتضم /5/ معياراً.

- كما هو مبين في الملحق (1) فإن معايير التقييم الخاصة بالوظائف القيادية والإشرافية مؤلفة من:
  - الكفاءات الأساسية, وتضم /12/ معياراً.
  - مهام الوظيفة, وتضم /5/ معايير.
  - أهداف الوظيفة وتضم /3/ معايير.

ب. المؤسسة العامة لسد الفرات[22]:

- كما هو مبين في الملحق رقم (2) فإن معايير التقييم الخاصة بالوظائف القيادية والإشرافية هي:
  - الكفاءات الأساسية, وتضم /8/ معايير.
  - مهام الوظيفة, وتضم /5/ معايير.
  - أهداف الوظيفة وتضم /4/ معايير.
- كما هو مبين في الملحق رقم (2) فإن معايير التقييم الخاصة بالوظائف الفنية والخدمية مؤلفة من:
  - الكفاءات الأساسية, وتضم /8/ معايير.
  - مهام الوظيفة, وتضم /5/ معايير.
  - أهداف الوظيفة وتضم /4/ معايير.

ج. النموذج المعمم من وزارة الشؤون الاجتماعية والعمل (نماذج أ, ب, ج)[23]

- كما هو مبين في الملحق رقم (3) فإن معايير التقييم الخاصة بالوظائف القيادية والإشرافية هي:

1. الكفاءات الأساسية, وتضم /16/ معياراً.

[22] اليونس, 2007, تقييم أداء العاملين والترفيعات السنوية, مرجع سبق ذكره, نموذج تقييم خاص بالمؤسسة العامة لسد الفرات.

[23] وزارة الشؤون الاجتماعية والعمل, لجنة تقييم أداء العاملين, نموذج تقييم الأداء المعمم بالأرقام (أ – ب – ج) لبعض الوزارات.

2. مهام الوظيفة, وتضم /4/ معايير.

3. أهداف الوظيفة وتضم /2/ معيار.

- كما هو مبين في الملحق رقم (3) فإن معايير التقييم الخاصة بالوظائف الفنية والخدمية تتألف من:

1. الكفاءات الأساسية, وتضم /15/ معياراً.

2. مهام الوظيفة, وتضم /2/ معيار.

3. أهداف الوظيفة وتضم /1/ معيار.

5.1.1. ملخص عملية تقييم الأداء في القطر العربي السوري

- بناءً على ما ذكر أعلاه حول طرائق وأساليب تقييم أداء العاملين في سورية ومن خلال تحليل نماذج واستمارات تقييم أداء العاملين التي تم جمعها فأنه يمكن ملاحظة أوجه التشابه التالية:
  - التقسيمات الأساسية للتقييم (الكفاءات الأساسية, مهام الوظيفة, أهداف الوظيفة).
  - العلامات المميزة للتقييم (0–2.49 و 2.5–3.49 و 3.5–5) والتي تحدد نتيجة التقييم ما بين ضعيف ووسط وجيد.
  - القائمين بتنفيذ عملية التقييم (الرئيس المباشر, المدير الأعلى, المدير الأعلى).
  - نماذج التقييم على اختلافها توسعت في تحديد معايير تقييم الكفاءات الأساسية ويلاحظ أن تعداد معايير هذا البند أكثر من تعداد معايير بندي مهام الوظيفة وأهداف الوظيفة على الرغم من أهميتهما.
- وجه الاختلاف الأساسي وذو الأثر الأكبر هو اختلاف معايير التقييم وتداخلها مع الملاحظة أنها متشابهة بالعموم من خلال التركيز على طريقة العمل والتنفيذ والأمن الصناعي والمكافحة الهدر وغيرها من المعايير إلا أنه لا يوجد معيار مقياسي لكل من الكفاءات الأساسية أو مهام الوظيفة أو أهداف الوظيفة وقد يتكرر المعيار الواحد ضمن الوظيفة الواحدة بعبارات مختلفة مما يؤدي لتداخل كبير في المعطيات.

2.1. عملية التفقد اليومي (Time & Attendance Process)

تعتبر الروح المعنوية للعاملين أحد الموضوعات الهامة التي يتابعها قسم الموارد البشرية والتي نالت القدر الكبير من العناية من جانب رجال الإدارة وعلماء النفس في الدول المتقدمة. ويرجع الاهتمام بموضوع الروح المعنوية لضمان فعالية أداء كافة العاملين في تحقيق أهداف المؤسسة وضمان التوافق والتعاون في ممارسة وظائفهم، وبالتالي تعاونهم مع الآخرين لتحقيق الكفاءة للأداء التنظيمي[24].

العديد من مؤشرات انخفاض الروح المعنوية مرتبط بمدى التزام العامل بأوقات الدوام الرسمي كارتفاع نسبة الغياب والتأخير, التباطؤ والتوقف عن العمل, التمارض ,الاستقالات, وغيرها من الشواهد والمظاهر تدل على القلق أو عدم استقرار الأداء، وهو ما يحتاج إلى دراسة وتعمق بشكل أكبر.

1.2.1. مقدمة (Introduction)

تعمد جميع المنظمات والمؤسسات للاستفادة من أنظمة التفقد اليومي الآلية منها أو اليدوية لتسجيل مدى التزام الموظفين بأوقات الدوام الرسمي والتأكد من تواجد العاملين في مواقع عملهم خلال ساعات العمل الرسمي وضبط تكاليف العمل من خلال خفض النفقات الإضافية وتجنب دفع الرواتب الوهمية. وبالتالي تعمد المؤسسات لحفظ جميع البيانات المتعلقة بالتزام الموظفين بمواعيد الدوام الرسمي والغياب المشروع كالإجازات الإدارية والخاصة والاستراحات والنقاهات أو الغياب غير المشروع والتأخر عن مواعيد الدوام الرسمي والتسرب ضمن ساعات العمل.

حسب (Mojca & Igor - 2011)[25] فإن بيانات أنظمة التفقد اليومي يمكن أن تشكل مؤشراً مهماً لعدد كبير من المشكلات سواء الشخصية منها أو تلك المتعلقة بشروط العمل. يمكن إرجاع المشكلات الشخصية التي تؤدي إلى تسرب العامل وعدم التزامه بساعات العمل الرسمي إلى المشكلات العائلية, ضعف الإمكانيات المادية

[24] المنتدى العربي لإدارة الموارد البشرية,

[25] Mojca B., Igor B., Knowledge Management and Information Technology in Analyzing Human Resource Processes, مرجع سابق, p.1

وعدم التنظيم والاختلال في الحياة الشخصية للموظف. أما المشكلات المتعلقة بشروط العمل فهي متنوعة كالمسافة الفاصلة ما بين مكان الإقامة ومكان العمل, عدم الرضا عن شروط العمل والشعور بالغبن أو ضعف تحفيز وتدريب الكوادر البشرية.

للتعمق أكثر في مدى التزام الكوادر البشرية بساعات العمل الرسمي ومعالجة مشكلات التسرب والغياب يجب تكوين نظرة شاملة لبيانات التفقد اليومي وتخطي النظرة الاحصائية إلى نظرة معرفية من خلال تحويل البيانات إلى معلومات ذات معنى ومن ثم تحويل هذه المعلومات إلى معرفة وربطها بمتغيرات العمل وفي النهاية مشاركة هذه المعرفة مع باقي القطاعات والفعاليات في المؤسسة.

إن التحدي الأكبر للتقنيات الحديثة لا سيما المعلوماتية منها هو إيجاد آليات لزيادة مستوى تحليل هذه البيانات بما يمنح المدراء وأصحاب القرار إمكانية الوصل للمعلومات بشكل أسرع وبصورة متكاملة بما يساعد في عملية اتخاذ القرار المناسب.

2.2.1. عملية التفقد اليومي في الجمهورية العربية السورية (Time & Attendance Process in S.A.R)

يتم إجراء عملية التفقد اليومي حسب القوانين والأنظمة النافذة في القطر العربي السوري والتي يمكن إيجازها بالنقاط التالية:

- تحديد عدد ساعات العمل بثماني ساعات يومية يجب أن يقضيها الموظف ضمن قطاع عمله أو خارجه بموجب مهمة تستدعي تواجده خارج هذا القطاع[26].
- عدد أيام الإجازات الإدارية المسموح بها خلال عام بشكل متناسب مع عدد سنوات الخدمة[27]:
  a. 15 يوماً لمن لا تقل خدمته الفعلية عن خمس سنوات.

[26] القانون الأساسي للعاملين في الدولة رقم /50/ لعام /2004/, المادة 42.

[27] المصدر ذاته, المادة 44.

b. 21 يوماً لمن بلغت خدمته الفعلية خمس سنوات وحتى عشر سنوات.

c. 26 يوماً لمن بلغت خدمته الفعلية عشر سنوات وحتى عشرين سنة.

d. 30 يوماً لمن أتم الخمسين من العمر أو تجاوزت خدمته الفعلية عشرين عاماً.

- منح قانون العمل إمكانية حصول الموظف على الإجازات المرضية بأنواعها المختلفة كالنقاهات والأمومة وغيرها ذلك وفق مدد محددة لكل منها.
- تجيز قوانين العمل في الجمهورية العربية السورية حصول الموظف على إجازة بدون راتب عند تجاوز المدة المحددة للإجازات الإدارية أو الاستراحات المرضية كالإجازات الخاصة الداخلية والخارجية أو الاستيداع أو الإجازات الدراسية وغيرها.
- يتم تسجيل مواعيد الحضور والانصراف بشكل يومي وتحفظ هذه المعلومات للمراجعة بشكل دوري أو عند الحاجة إليها وتعتبر العنصر الأهم عند حساب الأجور الشهرية.
- تتفاوت طرق جمع بيانات التفقد اليومي, بعض المؤسسات تعتمد الطرق الآلية كاستخدام أجهزة قراءة البصمة أو البطاقات الشخصية وغيرها بينما لاتزال مؤسسات أخرى تعتمد الطرق اليدوية كاستخدام التوقيع الشخصي لإثبات تواجد الموظف وفي كلتا الحالتين فإنه يتم تخزين وأرشفة البيانات على أجهزة الحاسب تمهيداً لاستخدامها بشكل إحصائي في عدة مجالات إدارية ومالية.
- تختلف نماذج تخزين بيانات التفقد اليومي حسب المؤسسات والمنظمات ولكن أغلبها يتضمن الحقول التالية:
  - الغياب المشروع: ويتضمن أسماء الموظفين موزعة ضمن حقول الإجازات بأنواعها (الإدارية, الخاصة, الساعية, الاستيداع وغيرها).
  - الغياب الصحي: ويتضمن أسماء الموظفين موزعة ضمن حقول النقاهات بأنواعها (الاستراحات المرضية, الأمومة, المشفى وغيرها).
  - المهمات: ويتضمن أسماء المكلفين بمهمات رسمية داخلية وخارجية.

- التأخر اليومي عن ساعة بدء الدوام الرسمي: ويتضمن أسماء المتأخرين في الوصول إلى مكان العمل ومقدار التأخير الحاصل.
- الغياب غير المشروع: ويتضمن أسماء الموظفين الذين لم يتواجدوا ضمن قطاع العمل دون وجود سبب موجه لذلك.

3.1 التنقيب في المعطيات (Data Mining)

التنقيب في المعطيات تكنولوجيا حديثة فرضت نفسها بقوة في عصر المعلوماتية في ظل التطور التكنولوجي الكبير وانتشار استخدام قواعد البيانات ويمكن تعريفها بأنها عملية البحث داخل مجموعة من البيانات عن معرفة غير مستكشفة, أي استكشاف المعرفة (Knowledge Discovery) من خلال استطلاع وتحليل كميات كبيرة من البيانات لاكتشاف علاقات وقواعد ذات معنى[28].

1.3.1. مقدمة (Introduction)

مع وجود كميات هائلة من البيانات المخزنة في قواعد ومستودعات البيانات ازدادت الحاجة إلى تطوير أدوات تمتاز بالقوة لتحليل البيانات واستخراج المعلومات والمعارف منها، من هنا ظهر ما يسمى بالتنقيب في المعطيات كتقنية تهدف لاستخراج المعرفة من كميات هائلة من البيانات.

إن التنقيب في المعطيات تقنية حديثة فرضت نفسها في مجال الذكاء الصنعي بما منحته للشركات والمنظمات من قدرة على استكشاف أهم المعلومات الموجودة في قواعد البيانات وفي جميع المجالات. كما تمنح تقنيات التنقيب القدرة على بناء التنبؤات المستقبلية واستكشاف السلوك والاتجاهات مما يسمح باتخاذ القرارات الصحيحة واتخاذها في الوقت المناسب وتجيب على العديد من الأسئلة، وفي وقت قياسي، بخاصة تلك النوعية من الأسئلة التي كان من الصعب الإجابة عليها، إن لم يكن مستحيلاً، باستخدام تقنيات الإحصاء الكلاسيكية، والتي كانت إن وجدت فإنها تستغرق وقتاً طويلاً والعديد من إجراءات التحليل.

وعليه فإن الكثير من الشركات والمنظمات الرائدة اليوم تستخدم من هذه التقنية المتقدمة لإنجاز عملية استكشاف المعرفة ضمن قواعد البيانات بشكل منهجي ومنظم بوصفها تشكل جوهر العمل الذي تعتمد عليه في

[28] أحسن, طيار, "التنقيب في البيانات واتخاذ القرارات", الملتقى الوطني السادس, جامعة سكيكدة, الجزائر, 27-28 كانون الثاني 2009.

تفعيل النشاطات وتحقيق الميزة التنافسية مع الأخذ بالحسبان بأن نجاح عمليات التنقيب في المعطيات ينبثق من الدمج ما بين الفهم الصحيح لطبيعة المعطيات مع تقنيات التحليل المتقدمة.

هذا ويقسم التنقيب في البيانات إلى نوعين رئيسين:

1. التنقيب الوصفي (Descriptive DM): ويعمل على تحليل البينات والروابط فيما بينها بغية استخراج العلاقات والروابط المخفية وبناء نموذج لم يكن موجود سابقاً اعتماداً على هذه البيانات والروابط.
2. التنقيب التنبؤي (Predictive DM) ويعمل على التنبؤ المستقبلي بآلية عمل البيانات بناء على نماذج مكتشفة في هذه البيانات أو من خلال نموذج تم تدريبه بياناته مسبقاً.

2.3.1. خوارزميات التصنيف (Classification Algorithms)

يمكن تعريف عمليات التصنيف (Classification) ضمن بيئة التعلم الآلي (Machine Learning) بأنها عملية توزيع البيانات – والتي تدعى بيانات التدريب – ضمن فئات مختلفة حسب خاصياتها المشتركة بالاعتماد على خوارزميات متعددة. تدرج عملية التصنيف ضمن عمليات التعلم المراقب (Supervised Learning) نظراً لوجود جميع القيم الحقيقية لجميع بيانات التدريب[29]. هذا وتعتبر عملية التصنيف أساساً لعملية التنبؤ (Prediction) من خلال النماذج التي يتم بناؤها أثناء عملية التصنيف والمرتبطة بنوع المصنف (Classifier) المستخدم وفيما يلي نستعرض المصنفات التي تم استخدامها ضمن هذا البحث والخوارزميات التي تستند إليها لإنجاز عمليات التصنيف المختلفة.

1) المصنف Naïve Bayes:

يستند هذا المصنف إلى نظرية بييز الاحتمالية (Bayes' theorem) القائمة على مبدأ الاحتمال الشرطي الذي يعمد لحساب احتمال وقوع أحد الأحداث الاحتمالية بناء على وقوع حدث آخر وفق المعادلة:

[29] لتمييزها عن التعلم غير المراقب (Unsupervised Learning) حيث لا تتواجد القيم الحقيقية أو فئات بيانات التدريب.

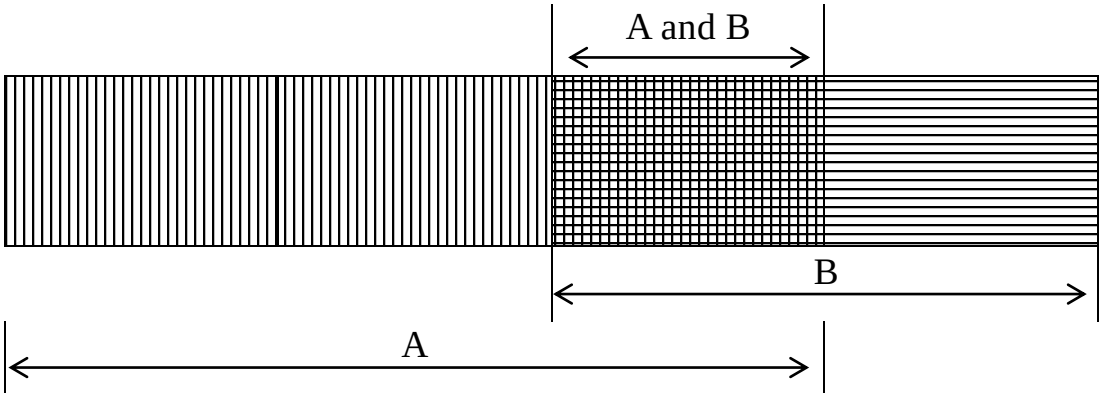
$$Prob(B \text{ given } A) = Prob(A \text{ and } B) / Prob(A)$$

Prob(B given A): احتمال وقوع الحدث B بناء على وقوع الحدث A

Prob(A and B): احتمال وقوع الحدثين A و B معاً أو ما يدعى (pairwise)

Prob(A): احتمال وقوع الحدث A أو ما يدعى (singleton)

ملاحظة: يبدو أن الحدث (B) حدث مرتبط بحدث مستقل وحيد (A) ولكن في الحقيقة فإنه يجب على هذه الخوارزمية التعامل مع ارتباط الحدث بعدة أحداث مستقلة.



الشكل 1.1 : آلية عمل الاحتمال الشرطي لنظرية بييز الاحتمالية.

يبين الشكل (1.1) تقاطع الاحتمالات حسب نظرية بييز وكمثال تطبيقي لتوضيح آلية عمل هذه الخوارزمية يمكن عرض المسألة التالية والمرتبطة بموضوع هذا البحث, ما هو احتمال أن يُسند للموظف وظيفة إشرافية وقيادية في حال كان تقييم أدائه بالنسبة لمكافحة الهدر والإسراف هو /4.5/؟ – وذلك بفرض وجود بيانات تدريبية مؤلفة من /100/ سجل تتضمن /75/ سجل ذات قيم أعلى من /4/ لمعيار مكافحة الهدر والإسراف منها /25/ سجل لعاملي الفئات الإشرافية القيادية.

الاحتمال المطلوب لهذه المسألة يمكن عرضه على الشكل التالي:

$$P(A \text{ and } B) = 25/100$$

$$P(A) = 75/100$$

$$\rightarrow P(B \text{ given } A) = 25/75 = 1/3$$

ومع تكرار العملية لكافة المتغيرات يتم بناء شبكة من الاحتمالات الشرطية التي تستخدم لبناء النموذج الخاص بهذه الخوارزمية والمرتبط بالعينة التدريبية.

ولتوضيح الملاحظة التي تم الإشارة إليها أعلاه بأن الحدث يمكن أن يرتبط بعدة أحداث مستقلة يمكن الإشارة هنا إلى الحالة التي يتم خلالها إسناد الوظيفة الإشرافية بناء على تقييم أدائه بالنسبة لمعيار مكافحة الهدر إضافة لمعايير أخرى كالتنظيم في العمل والدقة واتقان العمل وغيرها.

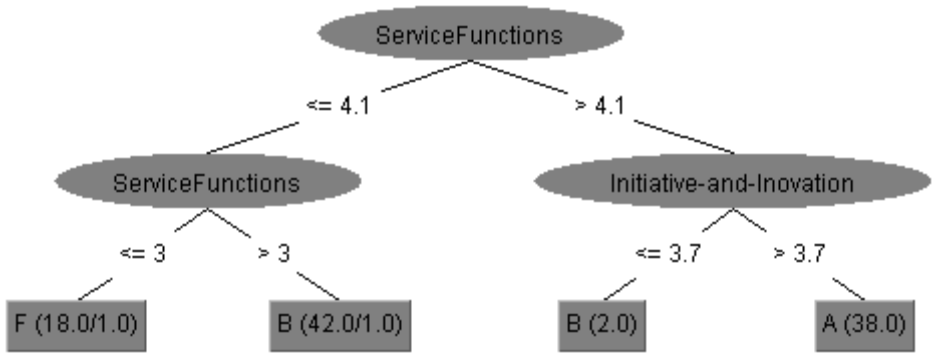
يمتاز هذا التصنيف بالسرعة في بناء النماذج كما أنه يمتاز بأنه قابل للتوسع (scalable) مع نمو بيانات التدريب وبتنفيذ عملية بناء النماذج بشكل متوازي (parallelized) ويمكن استخدامه لتصنيف بيانات ثنائية الفئات (binary class) أو متعددة الفئات (multi class).[30]

2) المصنف J48:

يندرج هذا المصنف ضمن خوارزميات أشجار القرار والتي على اختلاف أنواعها تشابه إلى حد ما خوارزمية تصنيف (Naive Bayes) من حيث اعتمادها على الاحتمالات الشرطية مع اختلاف رئيسي يكمن في أن هذه الخوارزمية تقوم بتوليد قواعد (Rules) لاستخدامها كجمل شرطية لتحديد السجلات والأحداث الاحتمالية بشكل عبارة شرطية (IF…..THEN).

يستند هذا النوع من التصنيفات إلى هيكلية شجرية مؤلفة من عقدة رئيسية تدعى الجذر (Root) ومجموعة عقد داخلية (Nodes) ومجموعة عقد نهائية (Terminals) بحيث يتضمن كل من الجذر ومجموعة العقد الداخلية القاعدة (Rule) التي تحدد المسار للفروع المرتبطة بما يسمح في النهاية بالوصول إلى النتيجة النهائية كما هو موضح في الشكل (2.1).

[30] Oracle Database On Line Documentation 11g Release (11.1), Data Warehousing and Business Intelligence, Data Mining, Classification, Naïve Bayes.



الشكل 2.1 : شجرة القرارات الناتجة عن التصنيف J48

حسب (Kumar, Sahoo - 2012)[31] تعمد هذه الخوارزمية إلى تقسيم مجموعة بيانات التدريب المراد تصنيفها إلى مجالات متقاطعة (Mutual Exclusive) ذات تسمية أو قيمة أو عملية لتوضيح وشرح البيانات داخل هذا المجال وذلك بالاعتماد على معيار يستخدم لحساب أو تعيين أفضل المعايير لتجزئة هذا المجال من البيانات التي يتم تدريبها والذي يدعى التابع الإحصائي (Information Gain) والمعرف بالمعادلة التالية:

$$Gain(S, A) = Entropy(S) - \sum_{v \epsilon Value(A)} \frac{|Sv|}{|S|} Entropy(s_v)$$

بحيث:

- مجموعة بيانات التدريب (S) ومجموعة المعايير (A).
- Values(A): جميع القيم الممكنة للمعيار A.
- S_v: مجموعة جزئية من المجموعة (S) المنتمية للمعيار (A) وذات قيمة (v).
- تابع العشوائية (Entropy): يعبر هذا التابع عن عشوائية المعطيات وتتراوح قيمته بين [0−1] ويعبر عنه بالمعادلة:

$$Entropy(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

[31] Kumar, Sahoo, 2012, Analysis of Bayes, Neural Network and Tree Classifier of Classification Technique in Data Mining using WEKA, Academy & Industry Research Collaboration Center, http://airccj.org/CSCP/vol2/csit2236.pdf

حيث يعبر المتغير (p_i) عن احتمالية انتماء مجموعة البيانات (S) إلى الفئة (i).

إن بناء الهيكلية الشجرية لأغلب أشجار التصنيف يتم من الأعلى إلى الأدنى (top-down) والاستفادة من طريقة البحث (greedy) لتحديد قيم التابع (Gain) لمختلف المعايير واختيار المعايير ذات القيمة الأعلى للتابع ($Gain$) ومن ثم إعادة العملية لباقي المعايير وصولاً لمجموعات جزئية متجانسة.

تتصف النماذج التي يولدها هذا النوع من التصنيفات بالدقة العالية والسرعة في بناء النموذج, كما يمكن تطبيقها على البيانات متعددة الفئات (multi class) وبأنها قابلة للتأويل والفهم من خلال تحليل شجرة القرار ومعاينة الرسم البياني المولد عند بناء النموذج[32] كما في الشكل (2.1).

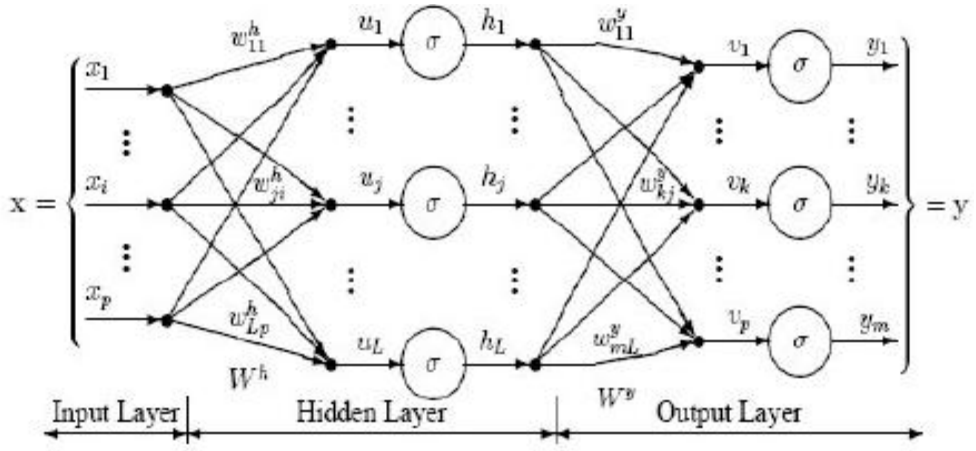
3) المصنف MLP (MultiLayerPerceptron):

يعتمد هذا التصنيف على خوارزميات الشبكة العصبونية (Artificial Neural Network) ذات طبقة أو أكثر بين الدخل والخرج بحيث ترتبط كل عقدة (node) في كل طبقة بجميع العقد الأخرى في باقي الطبقات, وباستثناء طبقة الدخل فإن جميع العقد هي عصبونات صنعية (Artificial Neuron).

تكمن الفكرة الأساسية لآلية عمل هذه الشبكة في خلق آلية عمل مشابهة للنظام العصبي العضوي المشكل من خلايا عصبية مترابطة تعمل مجتمعة لحل المشكلات التي تواجه الكائن الحي تمتاز بقدرتها على تعلم الاستجابة الصحيحة للمتغيرات المختلفة. ومع الاستفادة من القدرات البرمجية يمكن محاكاة عمل الخلية العصبية العضوية ومن ثم إيجاد شبكة من هذه العصبونات لتشكيل نظام عصبي صناعي متكامل.

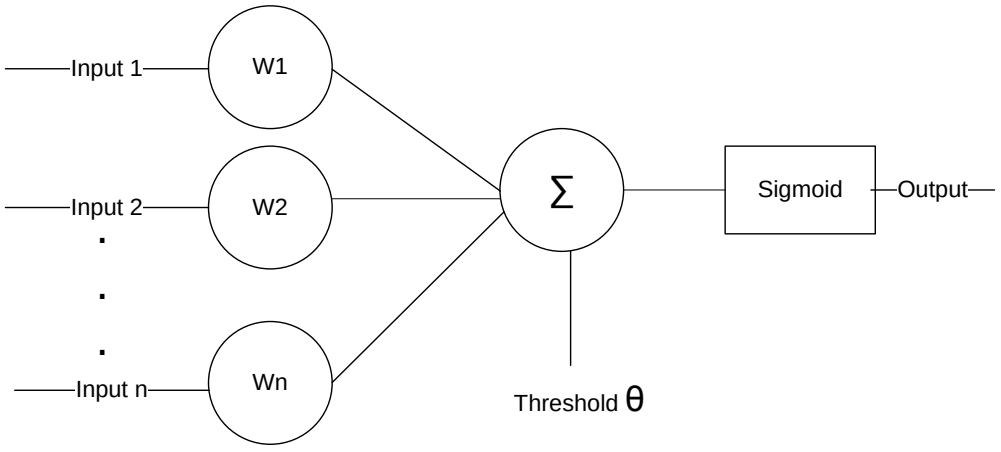
[32] Oracle Database On Line Documentation 11g Release (11.1), Data Warehousing and Business Intelligence, Data Mining, Classification, Decision Tree.

عند تدريب البيانات يتم إدخال البيانات عبر طبقة الإدخال (Input Layer) وتتم معالجتها ضمن الطبقات المخفية (Hidden Layers) وعرضها بالنهاية عبر طبقات الخرج (Output Layer).[33]



الشكل3.1 : الطبقات الثلاث لخوارزمية MultiLayerPerceptron

يظهر في الشكل (3.1) الطبقات الثلاثة المتعارف عليها في بنية الشبكات العصبونية حيث تتألف كل طبقة من واحدة أو أكثر من العصبونات الصنعية المتوازية, لكل عصبون كما يظهر في الشكل (4.1) عدد N من المدخلات ذات الوزن W لكل منها بالإضافة لمخرج واحد فقط. يقوم كل عصبون بدمج المدخلات مختلفة الأوزان من خلال جمعهم سوية وبالاستناد إلى حد العتبة θ يقوم بتحديد قيمة الخرج (Output).



الشكل 4.1 : بنية العصبون الصنعي الواحد

[33] Kumar, Sahoo, 2012, Analysis of Bayes, Neural Network and Tree Classifier of Classification Technique in Data Mining using WEKA, مرجع سابق, p. 361

لشرح آلية عمل هذه الخوارزمية بصورة مبسطة لابد من تعريف المتغيرات التالية:

- المدخلات $(x_1, x_2, \ldots, x_n)$ ذات الأوزان $(w_1, \ldots w_n)$.
- التابع u تابع يعبر عن احتمالية التنشيط (activation potential).
- تابع حد العتبة θ (threshold).
- تابع الخرج y (output).
- تابع التنشيط f (activation function)

وعليه يمكن تعريف تابع احتمالية التنشيط بالمعادلة:

$$u = \sum_{i=1}^{N} (w_i x_i) \qquad (1)$$

$$y = f(u - \theta) \qquad (2)$$

$$y = f\left(\sum_{i=1}^{N} (w_i x_i)\right) : w_0 = \theta, \quad x_0 = -1 \qquad (3)$$

تعبر المعادلة (3) عن الشرط الخاص بكل عصبون صنعي لجميع الطبقات مع اختلاف المتغيرات لدى الانتقال من طبقة إلى أخرى أو اختلاف الصف ضمن الطبقة الواحدة.

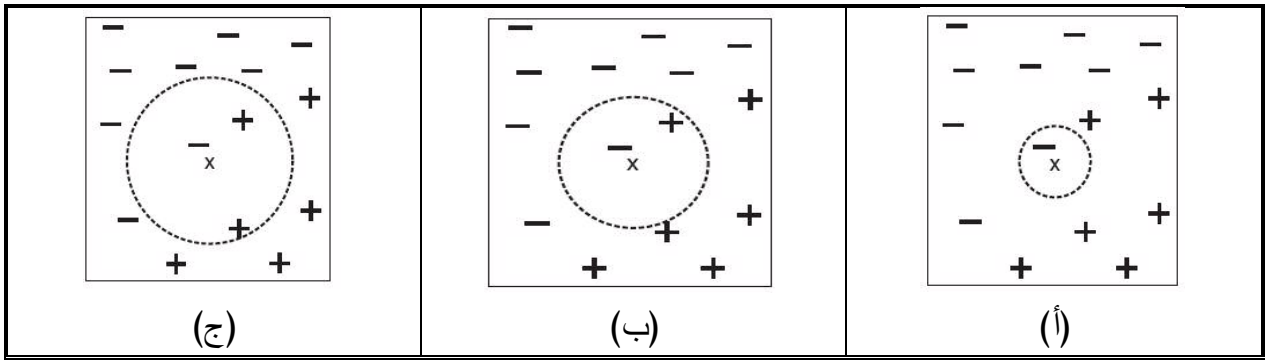
يتم تدريب شبكة البيانات بدايةً باختيار مجموعة أوزان بشكل عشوائي وحدود عتبات داخلية ومن ثم تعديل قيمة الأوزان بعد كل محاولة بشكل تكراري وصولاً للأوزان المناسبة بحيث تصبح ضمن حد مقبول.

إن قوة تصنيف هذه الخوارزمية غير الخطية مهّدت لاستخدامها بشكل واسع في عدة مجالات كتشخيص النطق (Speech Recognition), تشخيص الصور (Image Recognition) إضافة لبرامج الترجمة (Machine Translation) كما أثبتت هذه الخوارزمية قدرتها العالية في عمليات التخمين والتقريب (Approximation) ومعالجة الإشارات البيولوجية والاتصالات والالكترونية وصولاً إلى علوم الفضاء.

4) المصنف KNN:

يدرج هذا المصنف ضمن خوارزميات الجار الأقرب (K-Nearest Neighbor) التي تعتبر من التقنيات التنبؤية التي تهدف للتنبؤ بقيمة الفئة عن طريق مقارنة السجلات الشبيهة بالسجل المراد التنبؤ بقيمته وتقدير القيمة المجهولة لهذا السجل بناء على مقدار تلك السجلات. يعتمد عمل هذه الخوارزمية بشكل أساسي على عدد الحالات الأقرب (K) ووحدة القياس (metric) حيث يمثل الرمز (K) عدد الحالات المماثلة أي إيجاد مجموعة جزئية مؤلفة من (K) عنصر الأكثر تشابهاً مع الحالة المراد التنبؤ بقيمتها, وفي حال تماثل بعض أعضاء هذه المجموعة الجزئية يتم الاعتماد على وحدة القياس بغية تحديد العنصر الأكثر تشابهاً.

حسب (خلوف – 2010)[34] فإن الشكل رقم (1–5) يوضح آلية عمل هذه الخوارزمية حيث تظهر النقطة المجاورة الأقرب لإحدى نقاط البيانات المراد تصنيفها (X) ضمن الشكل (5.1.أ), بينما يظهر في الشكل (5.1.ب) النقطتين المجاورتين للنقطة (X), ويظهر الشكل (5.1.ج) النقاط الثلاثة المجاورة للنقطة (X).



الشكل 5.1 : مثال على توزع القيم عند استخدام التصنيف KNN

إن النقطة (X) في حالة الشكل (5.1.أ) تنتمي إلى الصف السالب, وفي حالة الشكل (5.1.ج) إلى الصف الموجب وذلك حسب نظام التصويت للأغلبية (Majority Voting Scheme), أما في حالة الشكل (5.1.ب) فإنه يتم اختيار الصف بناء على وحدة القياس (metric) ليتم تصنيف النقطة على

[34] خلوف, رزوق, شميس, تطوير آليات جديدة للتنقيب في المعطيات لإدارة علاقات الزبائن في بيئة مصرفية, مرجع سابق, ص

أساسه. يتم اختيار العدد (K) بشكل مناسب مع عدد البيانات بحيث يتم التغلب على التراكب الناتج عن عملية التصنيف والتي تزداد مع ازدياد شذوذ البيانات وعدم تناسقها.

5) التصنيف SVM (Support Victor Machine):

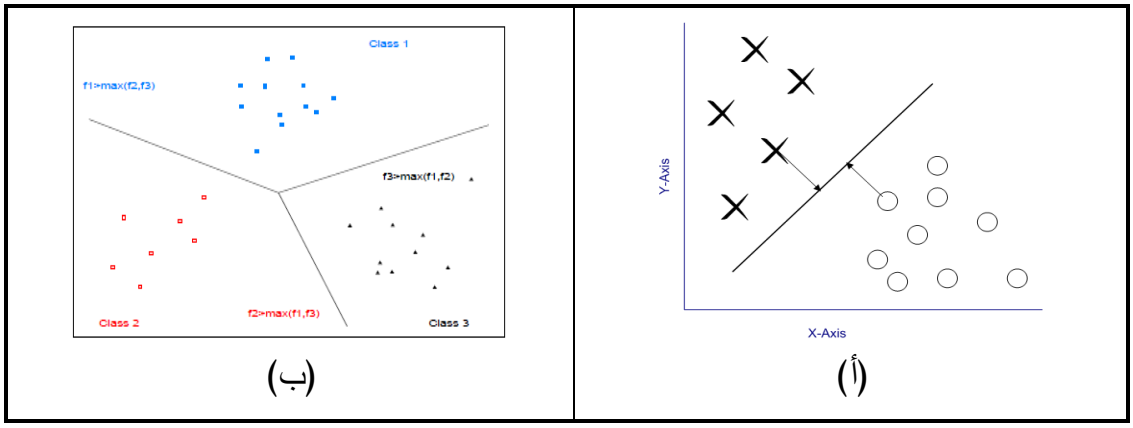
يعتبر هذا المصنف أحد أقوى المصنفات التقليدية لامتلاكه آلية عمل تدمج كلاً من خوارزمية الشبكات العصبونية مع خوارزمية الشعاع الأساسي (Radial Basis) لإيجاد أفضل سطح فاصل بين بيانات التدريب. يمتاز هذا المصنف بالمرونة, قابلية التوسع والسرعة في الأداء مما يعطيه الأفضلية في معالجة مسائل التشخيص (Recognition) المتنوعة وعلوم الأحياء (Bioinformatics), كما يتميز هذا المصنف بقدرته على معالجة المعطيات ذات عدد كبير من المعايير مقارنة بعدد سجلات البيانات المتواجدة.[11]

يعتمد مبدأ عمل هذا التصنيف على إيجاد أفضل سطح فاصل بين بيانات التدريب وفق حالتين:

1. تصنيف خطي: وذلك باختيار أفضل خط مستقيم يستطيع فصل البيانات ويكون أقرب ما يمكن لجميع هذه البيانات وهنا يمكن تمييز حالتين:[35]

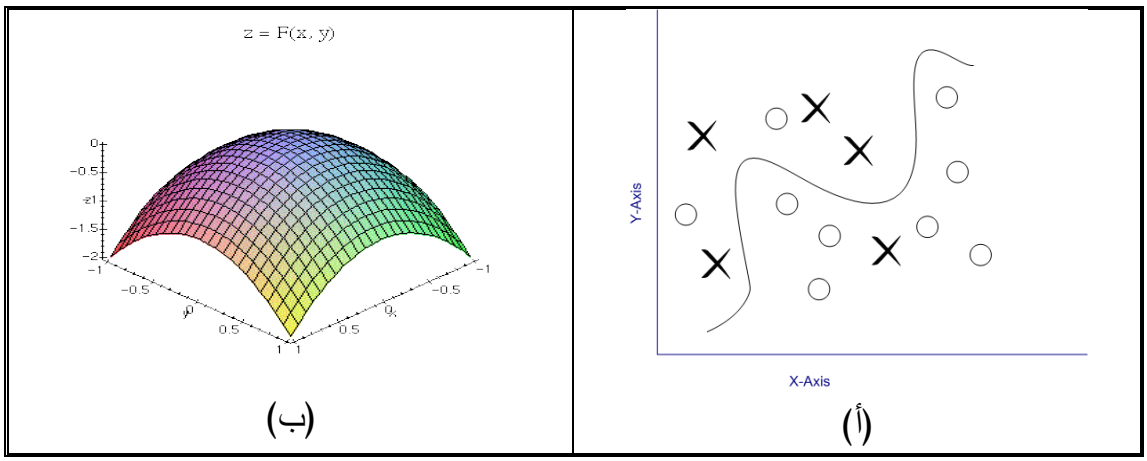
- مجموعة البيانات ثنائية الفئة (Binary Class) ذات بعدين (D=2), يبين الشكل (6.1.أ) مستقيم الحالة الأمثل (Optimum Situation) التي تقسم مجموعة البيانات إلى قسمين.
- مجموعة البيانات متعددة الفئات (Multi Class) متعددة الأبعاد (D>2), يوضح الشكل (6.1.ب) كيفية تقسيم مجموعة البيانات إلى ثلاثة صفوف من خلال ثلاثة خطوط مستقيمة تشكل الحالة الأمثل.

[35] www.wikipedia.org/Support_vector_machine.htm



الشكل 6.1 : أمثلة على توزع خطوط الفصل لتصنيف SVM الخطي

2. تصنيف غير خطي: وذلك باختيار أفضل سطح يستطيع فصل البيانات ويكون أقرب ما يمكن لجميع هذه البيانات وهنا يمكن تمييز حالتين حسب فئات وأبعاد البيانات فيما إذا كانت مؤلفة من ثنائية الفئة فقط (D=2) أو متعددة الفئات (D>2) كما يوضح ذلك الشكل (7.1.أ و 7.1.ب).



الشكل 7.1 : أمثلة على توزع خطوط الفصل لتصنيف SVM غير الخطي.

تم تطوير النسخة الأولية لهذا المصنف من قبل (Vladimir Vapnik) عام 1963 وفي عام 1995 تم تعميمها لمعالجة الحالات غير الخطية من قبل (Vapnik و Corinna Cortes).

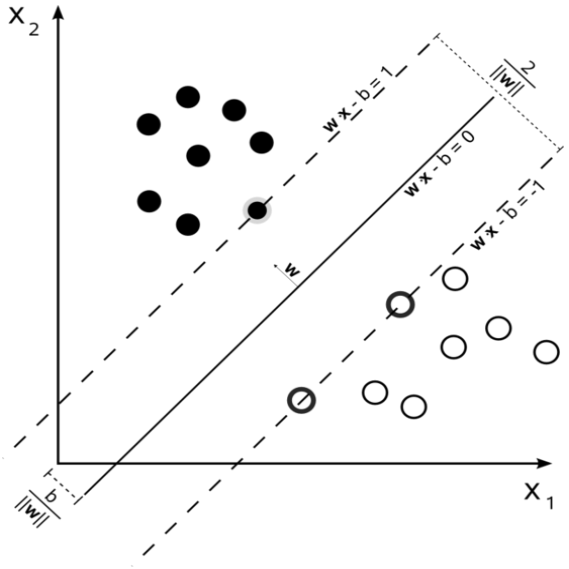
لتوضيح آلية عمل هذا المصنف في حالة التصنيف الخطي ثنائي البعد لمجموعة بيانات التدريب D المؤلفة من n عنصر والموضحة بالقانون:

$$D = \{x_i, y_i \mid x_i \in R^p, y_i \in \{-1, 1\}\}_{i=1}^{n}$$

حيث تتراوح قيم المتغير y بين القيمتين 1 أو 1− في حين x_i عبارة عن شعاع حقيقي عدد أبعاده p.

وبالتالي فإن الهدف هو إيجاد أفضل مستوي فاصل بأعلى مقدار من الهوامش (maximum-margin hyperplane) والقادر على فصل مجموعة النقاط ($y_i = 1$) عن مجموعة النقاط ($y_i = -1$), كما يظهر في الشكل (8.1) فإن الحد الفاصل المطلوب هو عبارة عن مجموعة النقاط x_i التي تحقق الشرط التالي:

$$w.x - b = 0$$



الشكل 8.1 : هوامش المستوي الفاصل بين بيانات التدريب

مع الأخذ بعين الاعتبار الجداء الشعاعي فإن w تعبر عن الشعاع العمود للسطح الفاصل وبالتالي لتعيين الخط الفاصل يستوجب إيجاد قيم كل من المتغيرين (w, b) بحيث تصبح المسافة بين الأسطح المتوازية أكبر ما يمكن. يمكن تعريف السطوح الفاصلة بالمعدلتين:

$$w.x - b = 1 \text{ and } w.x - b = -1$$

بفرض أن بيانات التدريب قابلة للفصل بشكل خطي فإنه بالإمكان إيجاد السطحين الفاصلين بحيث لا يتواجد أي نقاط مشتركة ومن ثم جعل المسافة بينهما أكبر ما يمكن.

باعتماد الطرق الهندسية فإن المسافة بين السطحين الفاصلين هي $(\frac{2}{||w||})$, وبالتالي فإن المطلوب هو أيجاد أدنى قيمة يمكن نسبها للمتغير w. ولمنع تواجد نقاط تدريبية داخل مجال الهوامش تم إضافة الشرطين التاليين:

$$w.x_i - b \geq 1 \qquad \text{or} \qquad w.x_i - b \leq 1$$

بدمج المعادلتين يصبح الشرط على الشكل التالي:

$$y_i(w.x_i - b) \geq 1 \text{ for all } 1 \leq i \leq n$$

تكمن صعوبة إيجاد الحلول للمعادلة بتعلقها بقيمة المتغير (||w||) ولكن باستبدال قيمة (||w||) بالمقدار المقابل لها رياضياً وهو ($\frac{1}{2}||w||^2$) وبالتالي فإن المطلوب هو تعديل المتغيرات حسب المعادلة التالية للحصول على أدنى قيمة ممكنة للمتغير w:

$$\arg\min_{(w,b)} \frac{1}{2}||w||^2$$

وبالأخذ بعين الاعتبار معادلة السطح الفاصل:

$$y_i(w.x_i - b) \geq 1$$

مع الاستفادة من معامل لاغرانج (Lagrange multipliers α) يمكن إعادة كتابة معادلة الحد الفاصل بالشكل:

$$\arg\min_{(w,b)} \max_{(\alpha \geq 0)} \{\frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i [w.x - b - 1]\}$$

وبالتالي يمكن حل المعادلة باعتبارها من الدرجة الثانية مع الأخذ بعين الاعتبار الشرط:

$$y_i(w.x_i - b) \geq 1$$

والنتيجة عبارة عن المتتالية

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

بالاستفادة من قيمة w بالإضافة للشرط $w.x_i - b = 1$ يمكن استنتاج قيمة المتغير b:

$$b = \frac{1}{N} \sum_{i=1}^{N} (wx_i - y_i)$$

3.3.1. خوارزميات العنقدة (Clustering Algorithms)

يمكن تعريف العنقدة بأنها عملية تقسيم مجموعة البيانات إلى مجموعات جزئية (Sub Sets) بحيث تكون بيانات المجموعة الواحدة ذات خصائص مشتركة ومتشابهة فيما بينها أكثر من بيانات المجموعات الأخرى. تعتبر عملية العنقدة أحد عمليات التنقيب الاستكشافي للمعطيات وأحد التقنيات الرائجة في التحليل الاحصائي للبيانات (Statistical Data Analysis), ويتم الاعتماد عليها في عدة مجالات منها التعلم الآلي (Machine Learning), تشخيص الأنماط (Pattern Recognition), تحليل الصور (Image Analysis) وعلوم الأحياء الحاسوبية (Bioinformatics).

إن عملية العنقدة بحد ذاتها ليست عبارة عن خوارزمية خاصة يمكن اجراؤها لمختلف أنواع المعطيات, وإنما عبارة عن عملية يمكن انجازها من خلال عدة خوارزميات متفاوتة بشكل كبير لاسيما في طريقة بناء المجموعة وتشكيلها ولا يمكن الفصل بشكل كامل بين أنواع وتقسيمات خوارزميات العنقدة نظراً لوجود تقاطع وتداخل فيما بينها ولكن يمكن تصنيف عمليات العنقدة حسب التقسيمات التالية:[36]

1. العنقدة القائمة على ارتباط المعطيات (Connectivity-based Clustering)

يطلق عليها أيضاً العنقدة الهرمية وهي قائمة على أن النقاط الاحصائية تكون أكثر ارتباطاً مع النقاط القريبة وتتناقص قوة الارتباط مع ازدياد بعد النقاط عن بعضها. إن خوارزميات هذا النوع من العنقدة تقوم بتشكيل المجموعة من خلال ربط نقاط المعطيات بعضها ببعض بالقدر الذي يسمح به الحد الأقصى من الطول المجاز (Maximum Distance) وبالتالي مع اختلاف هذه المسافة تختلف المجموعات التي تم تشكيلها. إن هذا النوع من الخوارزميات لا يمنح نمطاً واحداً لنقاط المعطيات وإنما بنية شجرية واسعة وممتدة من أجزاء المجموعات ومن ثم الدمج بينها بناءً على المسافة المسموح بها. إضافة إلى خيارات البعد والمسافة ما بين النقاط فإن هذه خوارزميات هذا النوع من العنقدة تتضمن خياراً مهماً آخر وهو معايير الرابط الذي يربط ما بين النقاط (Linkage Criterion) وذلك بالأخذ بعين

[36] www.wikipedia.org/ Cluster-analysis

الاعتبار أن المجموعة مؤلفة من نقاط متعددة وبالتالي يوجد عدة خيارات لحساب المسافات ولكن الخيارات الثلاثة الشائعة هي:

- العنقدة برابط وحيد (Single-Linkage Clustering): ويعتمد الحد الأدنى للمسافات.
- العنقدة برابط تام (Complete-Linkage Clustering): ويعتمد الحد الأعلى للمسافات.
- العنقدة برابط متوسط (Average-Linkage Clustering): ويعتمد متوسط المسافات.

يندرج تحت هذا النوع من العنقدة خوارزمية /COBWEB/ والتي يتم الاستفادة منها ضمن تطبيق WEKA. إن تعقيد هذه الخوارزميات هي من الدرجة $O(n^3)$ ويمكن أن تقليل هذه الدرجة $O(n^2)$ إلى عبر استخدام العنقدة برابط وحيد (Single-Linkage Clustering).

2. العنقدة القائمة على النقاط المركزية (Centroid-based Clustering):

يعمل هذا النوع من العنقدة إلى تقسيم نقاط المعطيات – أو المشاهدات الإحصائية – إلى عدد ثابت (K) من المجموعات وتعتبر خوارزمية Simple k-means هي المثال النموذجي والاكثر تطبيقاً لهذا النوع من العنقدة وأحد أشهر خوارزميات التعلم الآلي غير المراقب (Unsupervised Learning) حسب (Macqueen, 1967).

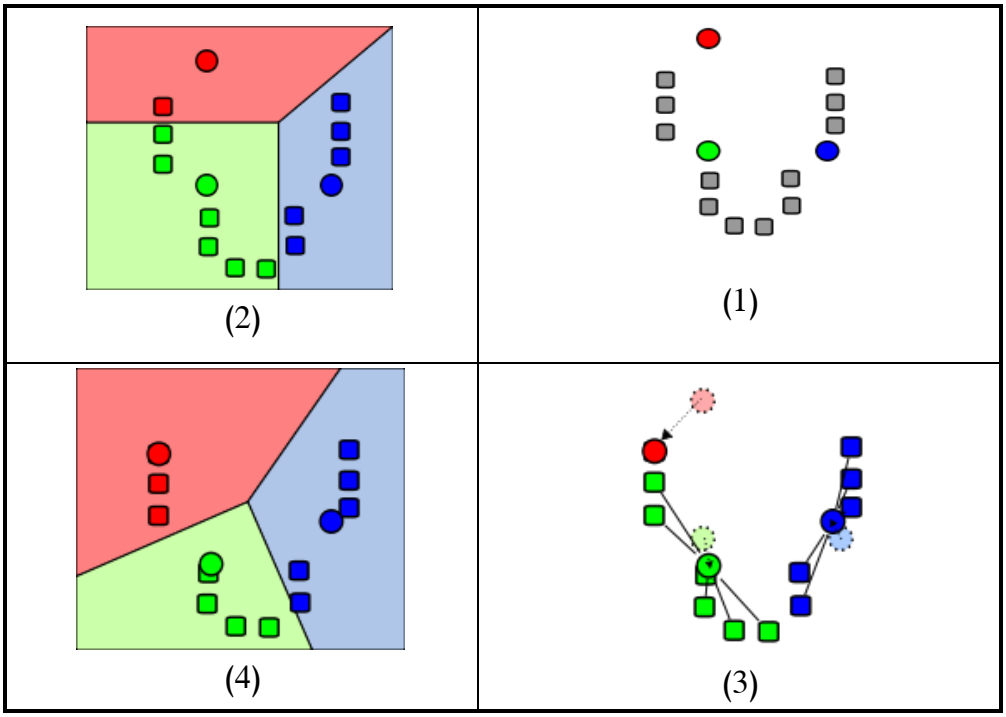
تتبع هذه الخوارزمية طريقة مبسطة في تقسيم المعطيات المتساوية في الأفضلية إلى مجموعات من خلال تعيين (k) مركز متوسط لتوزع المعطيات تمثل مراكز هذه المجموعات ومن ثم ربط ونسب جميع البيانات والنقاط إلى مجموعات حيث لا يسمح بوجود نقاط معلقة وغير مرتبطة. إن التوزيع الجيد لهذه المراكز ينتج عنه مجموعات صحيحة, والأفضل اختيار المراكز المتباعدة قدر الإمكان عن بعضها مما يقلل من حالات التداخل والتراكب بين المجموعات ويمكن تلخيص عمل هذه الخوارزمية بالخطوات الأربع التالية:

1. وضع (k) نقطة متوسطة (mean) ضمن فضاء البيانات المراد تجزئتها, وكما يظهر في الشكل (1.9.1) فقد تم توليد ثلاث نقاط بشكل عشوائي ضمن مجموعة البيانات المراد تقسيمها إلى ثلاث مجموعات.

2. نسب جميع نقاط المعطيات إلى المجموعة الأقرب مركزاً, كما هو مبين في الشكل (2.9.1) فقد تم نسب نقاط المعطيات إلى المراكز الثلاثة حسب قربها منها.

3. بعد توزيع جميع نقاط المعطيات ونسبها إلى المجموعات المختلفة يتم إعادة حساب نقاط مركزية للمجموعات بناء على نقاط المعطيات الموجودة في المجموعة بحيث تصبح النقاط المركزية الجديدة هي النقاط المتوسطة (mean) لكل من المجموعات الثلاث كما يظهر في الشكل (3.9.1) .

4. إعادة الخطوات /2/ و/3/ للوصول إلى الحالة المثلى التي تثبت خلالها النقاط المركزية وبالتالي فإن جميع البيانات موزعة على المجموعات بالحد الأدنى للمسافات التي أمكن حسابها كما يظهر في الشكل (4.9.1).



الشكل 9.1 : خطوات عمل خوارزمية العنقدة القائمة على النقاط المركزية.

بناء على الخطوات الأربع السابقة فإن درجة تعقيد هذه الخوارزمية هي (NP-Hard) وتعمل على إيجاد الحالة الأمثل المحلية (Local Optimum) نظراً لأن النتيجة النهائية هي الحالة المثلى ضمن شروط عدد المجموعات /k/ والتوزيع الابتدائي للنقاط المتوسطة.

ميزات استخدام هذه الخوارزمية:

1. مع وجود عدد كبير من البيانات واختيار عدد مقبول من المجموعات فإن هذه الخوارزمية تعتبر أسرع من العنقدة الهرمية.
2. إن المجموعات التي تنتجها هذه الخوارزمية تكون متراصة أكثر من المجموعات الناتجة عن العنقدة الهرمية ولا سيما في المجموعات الكروية.

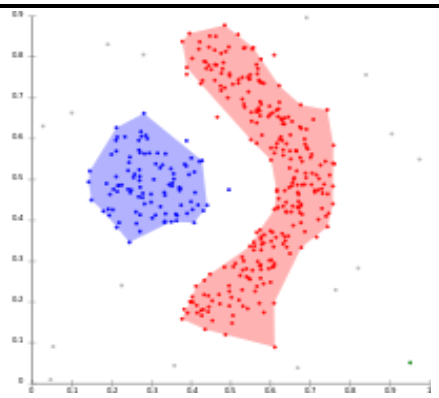
أما نقاط ضعف هذه الخوارزمية فهي:

1. صعوبة مقارنة مرجعية بناء المجموعات التي تم توليدها والعائدة إما للنقاط الابتدائية التي تم اختيارها أو لعدد المجموعات التي تم اختيارها مسبقاً.
2. العدد الثابت من المجموعات يجعل من الصعب التنبؤ بالعدد الأمثل للمجموعات التي من الممكن تقسيم البيانات إليها.
3. لا تعمل هذه الخوارزمية بشكل جيد في المجموعات غير الكروية.

3. العنقدة القائمة على كثافة المعطيات (Density-based Clustering):

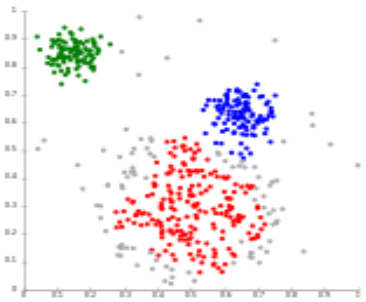
في هذا النوع من العنقدة تعرف المجموعات كنطاقات من نقاط المعطيات ذات كثافات متفاوتة لتوزع المعطيات وبالتالي فإن نقاط المعطيات المتناثرة يتم اعتبارها عادة كنقاط شاذة (noise), إن المبدأ الأساسي الذي يقوم عليه هذا النوع من العنقدة هو ربط نقاط المعطيات ضمن عتبات مسافات وأبعاد معينة وذلك بشكل مشابه للعنقدة الهرمية إلا أنه يربط النقاط التي تلبي معيار الكثافة (Density Criterion) الذي يعرف بأنه الحد الأدنى لعدد نقاط البيانات ضمن نصف القطر وبالتالي فإن نقطة التفوق الرئيسية لهذا النوع من

العنقدة على العنقدة الهرمية هو انخفاض تعقيد خوارزميات هذه النوع من العنقدة إلى الدرجة الخطية بينما تتمثل نقطة الضعف الرئيسية لهذا النوع في استكشاف حدود المجموعات حيث تتوقع خوارزميات هذا النوع انخفاض الكثافة للكشف عن حدود المجموعة وهو ما قد لا يكون صحيحاً في العديد من الحالات مما يسبب مشكلة في فصل المجموعات المتقاربة, كما يظهر في الشكل (10.1) فإن العنقدة باستخدام خوارزمية DBSCAN تختلف مع اختلاف كثافة المعطيات على حدود المجموعات بينما تكون النتائج أفضل باستخدام خوارزمية OPTICS.



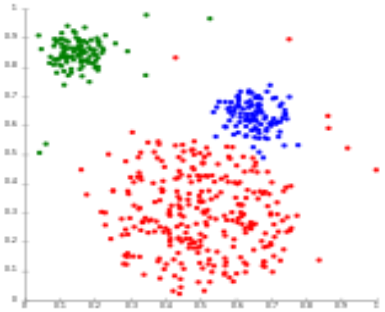
(1)

العنقدة باستخدام DBSCAN – حالة انخفاض الكثافة عند حدود المجموعات.



(2)

العنقدة باستخدام DBSCAN – حالة عدم انخفاض الكثافة عند حدود المجموعات.



(3)

العنقدة باستخدام OPTICS – حالة عدم انخفاض الكثافة عند حدود المجموعات.

الشكل 10.1 : الحالات المختلفة لعمل خوارزميات العنقدة القائمة على كثافة المعطيات.

أشهر خوارزميات هذه النوع من العنقدة هما خوارزمية DBSCAN وخوارزمية OPTICS, حيث تعتبر خوارزمية OPTICS تعميم لخوارزمية DBSCAN التي تضم المجموعات المشكلة خلال استخدامها جميع نقاط المعطيات التي تلبي شرط الكثافة بالإضافة إلى نقاط المعطيات التي تدخل

ضمن نطاق هذه النقاط والتي يتم تحديدها من خلال متغير خاص يمكن تحديده مسبقاً وقد تم حذف هذا المتغير في خوارزمية OPTICS وتم الاستعاضة عنه بإدراج نقاط المعطيات التي لا تحقق شرط الكثافة في بنية الهرمية R-tree.

4.3.1. انتقاء المعايير (Attribute Selection)

يمكن تعريف عملية انتقاء المعايير بأنها عملية اختيار أصغر مجموعة ضمنية من المجموعة الأصلية التي تضم جميع المعايير. تعتبر هذه العملية معقدة وأغلب الخوارزميات التي تعمل ضمن هذا المجال ذات درجة NP-hard كما تعتبر أحد المجالات الفعالية في علوم الحاسب وخاصة في أبحاث تشخيص الأنماط (Pattern Recognition) والتعلم الآلي والتنقيب في المعطيات.

إن عملية انتقاء المعايير هي عملية جوهرية في بعض المسائل وخاصة عمليات التوقع (Prediction) وتشخيص الأغراض (Object Recognition) والنمذجة (Modeling) – ربما تكون جميع المعايير مهمة في بعض الحالات ولكن من أجل بعض الأهداف يكتسب بعضها أهمية خاصة.

إن عمليات انتقاء المعايير تشارك بشكل مباشر في زيادة سرعة تنفيذ خوارزميات التصنيف ورفع أداء التنقيب في المعطيات وذلك من خلال تخفيض أبعاد المعايير وحذف البيانات المتكررة والغير مترابطة أو ذات القيم الشاذة عن باقي قيم المجموعة وبالتالي تعتبر أحد العمليات الرئيسية المستخدمة في عمليات التصنيف ويوجد عدة خوارزميات للقيام بهذا العمل نذكر منها:

1. خوارزميات الترتيب (Ranker): تعتبر أبسط خوارزميات انتقاء المعايير من خلال ترتيب المعايير بناء على التقييم الخاص بكل معيار على حدا ومن ثم توليد قائمة لجميع المعايير مرتبة حسب درجة تقييمها.[37] تستخدم هذه الخوارزمية مع عدد من مقيمات المعايير (Attribute Evaluators) نذكر منها (Chisquare, GainRatio, InfoGainetc).

[37] Megha A., (2013), Performance Analysis Of Different Feature Selection Methods In Intrusion Detection, INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, 2,225-231

2. خوارزمية (GreedyStepwise)[38] التي تستفيد من خوارزمية Greedy ضمن فضاء المعايير الموجودة, حيث تقوم باختيار نقطة عشوائية من مجموعة البيانات وتبدأ بإضافة أو حذف نقاط أخرى من فضاء البيانات للوصول إلى النتيجة الأفضل, ومن ثم تقوم بترتيب المعايير التي تم اختيارها حسب أهميتها وبذلك تؤمن الحالة الأمثل محلياً (Local Optimum).
3. خوارزمية (Genetic Search)[39]: تعتمد خوارزمية الجينات المبسطة لإيجاد الحالة الأمثل (Optimal Solution) ضمن فضاء البيانات. تستخدم هذه الخوارزمية مع عدد من مقيمات المعايير (Attribute Evaluators) نذكر منها (CFSSubEval, ConsistencySubsetEval, WrapperSubsetEval).

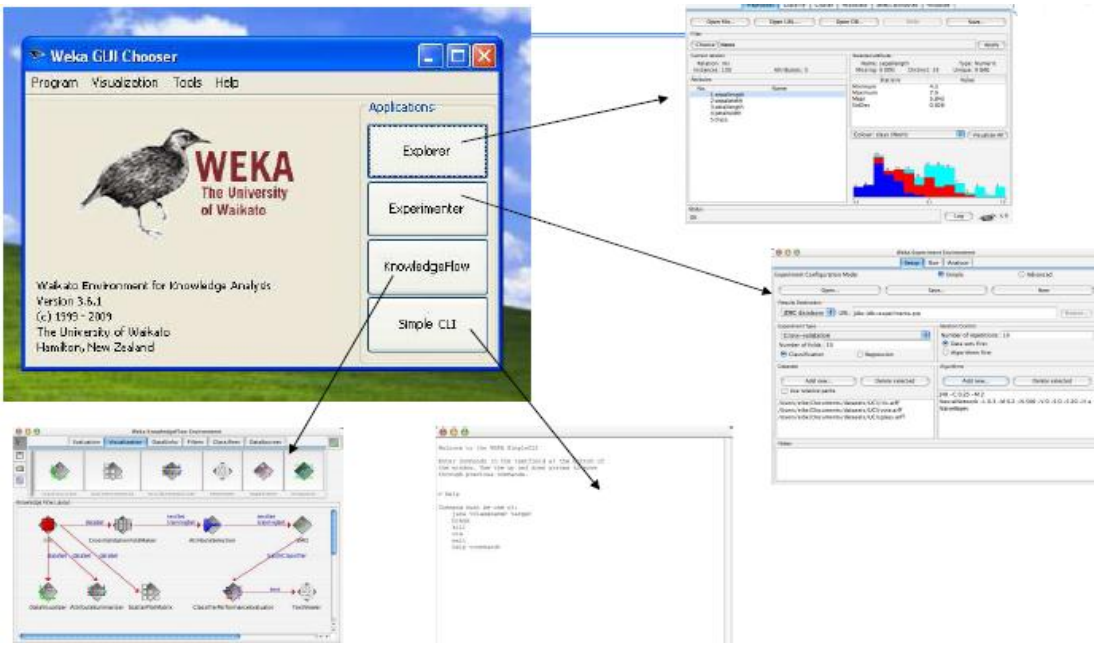
---

[38] المصدر ذاته.

[39] المصدر ذاته.

4.1. تطبيق WEKA (Waikato Environment for Knowledge Analysis)

تطبيق حديث ومتجدد يؤمن إجراء عمليات التنقيب في المعطيات بسهولة من خلال تزويده بمجموعة ضخمة من خوارزميات التعلم الآلي والعديد من الميزات التي تجعله أحد أهم التطبيقات المعروفة للتنقيب في المعطيات في الوقت الحاضر. الشكل (11.1) يبين الواجهة الرئيسية لهذا التطبيق.



الشكل11.1 : صورة توضيحية لواجهة عمل تطبيق WEKA.

1.4.1. مقدمة (Introduction)[40]

تم تطوير تطبيق WEKA في مخابر جامعة (Waikato) في نيوزيلندا بالاعتماد على لغة البرمجة (JAVA) وكانت النسخة الأولى عام 1996, ويعتبر أحد أفضل برمجيات التنقيب في المعطيات المعروفة حتى الآن حيث يعمل بالتوافق مع كافة أنظمة التشغيل المعروفة (Widows, Linux, Macintosh), ويمكن الوصول والتعامل مع كافة عناصر هذا التطبيق من خلال لغة البرمجة جافا كما يمكن تحميل البيانات إلى التطبيق من عدة مصادر تتضمن:

[40] الموقع الرسمي لتطبيق WEKA, http://www.cs.waikato.ac.nz/ml/weka

1. الملفات ذات الصيغ المعرفة لهذا التطبيق (arff, xrff, data, csv, names, bsi)

2. مختلف قواعد البيانات (SqlServer, MySql, Oracle, …).

3. مسارات ونطاقات عناوين الحواسب ضمن الشبكات الحاسوبية (URLs).

يضم هذا التطبيق مجموعة كبيرة من أدوات التنقيب في المعطيات والتعلم الآلي مجانية المصدر وتتضمن الواجهة الرسومية العمليات الرئيسية التالية:

- تحضير البيانات (Pre–processing Data)
- التصنيف (Classification)
- التجزئة (Clustering)
- قواعد الربط (Association Rules)
- اختيار المعايير (Select Attributes)

2.4.1. الميزات الرئيسية (Main Features)

- أدوات لتحضير وتجهيز البيانات (Preprocessing Tools) عدد /49/.
- خوارزميات تصنيف (Classification Algorithms) عدد /76/.
- خوارزميات تجزئة (Clustering Algorithms) عدد /8/.
- مقاييس للمعايير (Attribute Evaluators) عدد /15/.
- خوارزميات انتقاء أهم المعايير (Attribute Selection) عدد /10/.
- خوارزميات لإيجاد قواعد الربط (Association Rules) عدد /3/.
- واجهة إدخال غير رسومية عدد /1/, واجهة رسومية (UI) عدد /3/ وهي:

- المستكشف (The Explorer): ويضم كلاً من تحضير البيانات (Preprocess), التصنيف (Classification), التجزئة (Clustering), التجميع (Association), اختيار المعايير (Attribute Selection), التمثيل البياني (Visualization).
- المختبر (The Experimenter): يتضمن عمليات اختبار وتقييم خوارزميات التعلم الآلي.
- تدفق المعرفة (The Knowledge Flow): ويتضمن إمكانية التصميم المرئي لعمليات استكشاف المعرفة
- واجهة غير رسومية بسيطة: وهي واجهة بسيطة يمكن من خلالها كتابة الأوامر التنفيذية.

3.4.1. المصطلحات الرئيسية (Basic Terms)

فيما يلي تعريف وتوضيح لبعض للمصطلحات التي سيتم استخدامها والمرتبطة بتطبيق WEKA, علماً أن أغلب هذه المصطلحات ذات تعاريف رياضية وإحصائية منفصلة عن هذا التطبيق ولكن ضمن هذه الدراسة نعتمد على تعريفها وبيان أهميتها ضمن التطبيق المستخدم.

1.3.4.1. مصفوفة الشك (Confusion Matrix)

مصفوفة الشك أو كما تدعى أيضاً مصفوف الخطأ (Error Matrix) هي عبارة عن جدول يسمح بتشخيص مدى أداء خوارزمية التصنيف من خلال توزع الصفوف داخل هذه المصفوفة بحيث يمثل العمود النتيجة المتوقعة ويمثل السطر النتيجة الحقيقية كما في الشكل (12.1).

		التصنيف المتوقع		
		A	B	C
التصنيف الحقيقي	*A*	37	2	0
	B	4	40	0
	C	0	4	13

الشكل 12.1 : مصفوفة الشك الخاصة بتطبيق WEKA

إن قطر المصفوفة يمثل القيم المتوقعة بشكل صحيح ومطابقة للواقع فالخلية ([A, A] = 37) تمثل عدد الحالات التي توقع خلالها التصنيف القيمة (A) متغير قيمته الأصلية هي (A), أما في الحالة ([A, B] = 2) فإن التصنيف توقع القيمة (A) لمتغير قيمته الأصلية هي (B), وبالتالي فإن دقة التصنيف ناتجة من مجموع قيم القطر بالنسبة للعدد الكلي وفي هذه الحالة هي ($\frac{90}{100}$), وبالتالي فإن ازدياد معطيات القطر وقلة المعطيات المحيطة بالقطر دليل على قوة أداء التصنيف وصولاً إلى الحالة المثالية التي تصبح فيها دقة التصنيف مساوية ($\frac{100}{100}$).

2.3.4.1. معامل الإحصاء كابا (Kappa Static)

يعتبر هذا المعامل أحد أهم المؤشرات لقوة التصنيف من خلال قراءته لمصفوفة الشك واختصار دقة التصنيف الموجود داخلها إلى رقم وحيد يتراوح ما بين (0 إلى 1) من خلال تطبيق المعادلة التالية:

$$k = \frac{P(a) - P(e)}{1 - P(e)}$$

P(a): نسبة عدد حالات التطابق.

P(e): نسبة عدد الحالات التصادفية.

كمثال من الجدول رقم (12.1) يمكن حساب قيمة معامل الإحصاء كما يلي:

- P(a) = (37 + 40 + 13) / 100 = 0.9
- P(A) = $(\frac{37+4}{100})$ × $(\frac{37+2}{100})$ = 0.1599
- P(B) = $(\frac{40+4}{100})$ × $(\frac{4+40+2}{100})$ = 0.2024
- P(C) = $(\frac{13+4}{100})$ × $(\frac{13}{100})$ = 0.0221
- P(A) = P(A) + P(B) + P(C) = 0.3844
- k = 0.376

4.4.1. الحزمة البرمجية

إحدى أهم ميزات التطبيق والتي لا تزال قيد التطوير هي تزويده بحزمة (Package) تضم أغلب تقنيات التنقيب في المعطيات للتعامل المباشر مع الشيفرة المصدرية (code) المكتوبة بلغة البرمجة جافا. وبالتالي يمكن تضمين عمليات التنقيب في المعطيات ضمن البرمجيات التي يتم تطويرها دون الحاجة للاستعانة بتطبيق (WEKA) بشكل منفصل.

يؤمن الموقع الخاص بالتطبيق إمكانية تحميل الحزمة الخاصة بهذه الإمكانية والتي تدعى (weka.jar), والتي يتم إضافتها إلى مجموعة حزم التطبيق المراد تنفيذه, تضم الحزمة العناصر التالية:

- *Instances*: لتضمين البيانات والمعطيات المراد الاستفادة منها.
- *Filter*: للمعالجة المسبقة للبيانات قبل تطبيق تقنيات التنقيب.
- *Classifier/Clusterer*: لمعالجة البيانات.
- *Evaluating*: لتحديد مدى جودة العمليات المنفذة.
- *Attribute Selection*: لانتقاء أهم المعايير المستخدمة.

يضم الملحق رقم (4) مثال تطبيقي للاستفادة من هذه الحزمة ضمن بيئة العمل البرمجية لتنفيذ بعض العمليات المستخدمة ضمن هذا البحث.

4.4.1. الميزات والفوائد

- يندرج هذا التطبيق ضمن تصنيف البرمجيات مجانية المصدر (Open Source) وبالتالي يمكن للمحترفين الاستفادة من هذه الخاصية للتعديل على الشيفرة البرمجية بما يتناسب والعمل المراد تنفيذه.
- يتضمن بيئة عمل مرئية (GUI) ذات مرونة عالية جداً تتيح اجراء جميع العمليات ومشاهدة النتائج من خلال الرسوم والأشكال البيانية.
- توافق التطبيق مع أغلب أنظمة التشغيل وتضمين عناصره ضمن حزمات برمجية للاستفادة منها ضمن تطبيقات ومحررات برمجية (Editors) ومنها Matlab, Eclipse وغيرها.

5.4.1. نقاط الضعف

يمثل ضعف التوثيق وقلة المصادر الموثوقة والمناسبة أحد أهم مساوئ هذا التطبيق, وعلى الرغم من وجود مكتبة موجودة ضمن الموقع الخاص بهذا التطبيق والتي تتضمن مستندات متعددة من بينها دليل استخدام موسع إلا أن النشر في مجال استخدام هذا التطبيق محدود ولا سيما في معالجة المشكلات التي تظهر أثناء إجراء البرنامج وفي طرح أمثلة عملية تمثل إجراء العمليات المتعددة المتضمنة داخل التطبيق وخاصة فيما يتعلق بالحزمة البرمجية الملحقة.

الفصـــــــل الثـــاني

المنهــج العمـلي

(Practical Method)

تمهيد:

يتناول هذا الفصل عرض الخطوات العملية التي تم تنفيذها لإنجاز هذا البحث بدءاً من جمع المعطيات حسب النماذج المقترحة ومعالجتها عبر تطبيق WEKA وصولاً لبناء النموذج المقترح.

1.2 عينة البحث (A Research Data Set)

تم جمع المعطيات التي سيتم استخدامها في هذا البحث من خلال بناء نموذج موحد لجمع نتائج تقييم أداء العاملين وجدولتها تمهيداً لاستخدامها, ومن ثم إرفاق بعض المعلومات الأساسية للعاملين إلى بيانات التفقد اليومي الخاصة بهم وجدولتها تمهيداً لتحليلها.

2.1.2. النموذج الموحد لتقييم الأداء (Unified Form for Performance Evaluation)

لتأمين القدرة على استخدام بيانات التقييم ضمن بيئة عمل التنقيب في المعطيات لابد من إيجاد هيكلية موحدة ومناسبة للمعايير المستخدمة في التقييم وبالتالي لابد من وجود نموذج موحد يضم جميع المعايير بشكل غير متداخل وذلك بالاعتماد على أوجه التشابه والاختلاف التي تم توضيحها في الفصل السابق. يبين الجدول (1.2) النموذج الموحد لتقييم أداء العاملين ويمكن ملاحظة ما يلي:

- تم استثناء فئة الوظائف الإنتاجية من هذا البحث كون الترتيب الوظيفي الخاص بهذه الفئة لا يسمح لها بالتنقل ضمن الفئتين الأخريين بينما إمكانية التنقل بين فئة الوظائف الإشرافية والقيادية وفئة الوظائف الفنية والخدمية ممكن حسب ما تم توضيحه في الفصل الأول من هذا البحث.
- تم إعطاء الوظائف الإشرافية والقيادية الترميز /الوظيفة A/, الوظيفة الفنية والخدمية الترميز /الوظيفة B/.
- تم إضافة /الوظيفة F/ والتي ترمز إلى الفئة التي يجب الاستغناء عن خدماتها وذلك لعدم كفاءتها.
- تم فصل معايير مهام الوظيفة والأهداف ودمج المعايير الخاصة بها بمعيار واحد فقط لكل منها بحيث يمكن تعريفها كما يلي:

- مهام الوظيفة: وتشمل الالتزام بتنفيذ جميع الأعمال التي من شأنها تسهيل عمل المؤسسة في سبيل تحقيق أهدافها.
- الأهداف: وتتضمن الإيمان بأهداف المؤسسة والعمل بتفان لتحقيق هذه الأهداف وتنفي الخطط والتعليمات الموضوعة في سبيل ذلك.

الجدول 1.2 : النموذج الموحد لتقييم أداء العاملين.

معيار التقييم Evaluation Attribute	التوضيح Description
الكفاءات الأساسية (Basic Skills)	
المعرفة Knowledge	• فهم طبيعة العمل والمتطلبات والأدوات المرتبطة بهذا الموقع من العمل. • امتلاك القدرة العلمية والكفاءة اللازمة للقيام بواجبات هذا الموقع من العمل.
التنفيذ Execution	• أداء العمل ضمن الوقت المحدد وبأقل مستوى أخطاء ممكن. • المتابعة باستمرار إنجاز المهام المطلوبة مع المرؤوسين. • أداء المهام ضمن الوقت المحدد وبأقل مستوى من الأخطاء.
التعرف على المشكلات وإيجاد الحلول Problem Solving and Resolution	• القدرة على إيجاد الحلول للمشكلات في الوقت المناسب. • التعرف على مشكلات العمل واتخاذ القرارات المناسبة وتحمل المسؤولية بشكل فعال. • حل مشكلات المراجعين, تحديد مشكلات المراجعين بشكل واضح والسعي لحل مشكلاتهم والإجابة على استفساراتهم.
التعليم الذاتي وتطوير الذات Self−Development	• امتلاك القدرة على التعامل مع تقنيات الإدارة الحديثة. • الاهتمام بتنمية قدراته الذاتية.
السلامة Safety	• الاهتمام بمتابعة قضايا الصحة والسلامة المهنية ومشكلاتها. • التقيد بإجراءات وبرامج الصحة والسلامة المهنية. • الالتزام بقواعد الأمن الصناعي والصحة والسلامة المهنية. • التدرب على أساليب العمل الآمن وحث الآخرين على ذلك إضافة إلى تعريف طرق لتحسين السلامة ضمن بيئة العمل.
الإنتاجية Productivity	• كمية العمل المنجزة بالمقارنة مع الأقران.
الجودة Quality	• جودة العمل المنجز بالمقارنة مع الأقران.
المبادرة والإبداع Initiative-and-Innovation	• القدرة على المبادرة والابتكار والابداع. • المبادرة في البدء بإنجاز الأعمال وإنجازه قبل البدء التوجيهات التعليمات .
المواظبة والدقة Attendance -and-precision	• التقيد بفترة الدوام الرسمي والعمل حسب الجدول اليومي. • المواظبة على الدوام وأوقات الحضور والانصراف.
التنظيم Organization	• تنظيم مكان العمل بما يناسب طبيعة العمل. • تنظيم أوقات العمل وأولوياته بما يحقق انجاز الأعمال دون تراكمات.
قابلية التكيف Adaptability	• القدرة على التكيف مع التغييرات في محيط العمل والمتطلبات والاولويات. • القدرة على التركيز بالعمل في ظروف ضغوط العمل.
الأخلاق والتهذيب Politeness	• التعامل مع الناس باحترام ولطف ويقدم الإرشاد ويدعم الآخرين. • التعامل مع المراجعين بلطف واحترام وبشكل واضح ومفهوم. • الإيمان بأن العميل على حق دائماً وإظهار الأدب للمراجعين والزبائن.

المظهر الشخصي Personal Appearance	• ارتداء اللباس والهندام المناسب عند اللقاء مع المراجعين والزبائن , والهندام المناسب للمحيط والذي يعطي الانطباع المناسب لدى الزبائن والمراجعين. • المحافظة على حسن المظهر ويلتزم باللباس الرسمي في حال وجوده.
الاقتصادية Economic	• الاستخدام الاقتصادي لمستلزمات العمل المادية. • التخطيط والعمل بشكل فعال وعملي.
التفكير بشكل استراتيجي Strategic Thinking	• القدرة على التنبؤ والتخطيط وجدولة العمل. • القدرة على استشراف المستقبل والتنبؤ به.
العلاقة مع شركاء العمل Interaction with Coworkers	• العمل بشكل جيد مع زملاء العمل, والاحترام المتبادل مع الآخرين • التشجيع على التعاون وروح الفريق. • التعاون مع الآخرين لإنجاز المهام المشتركة ويمتلك روح الفريق بالعمل.
العلاقة مع المشرفين والرؤساء Interaction with Supervisors	• العمل بشكل جيد مع الرؤساء واحترام سلطتهم والتعامل معهم بشكل احترافي.
العلاقة مع الزبائن/المرؤوسين Interaction with Clients/Clients	• العمل الجيد مع الزبائن والمراجعين وتشكيل العلاقات الجيدة معهم والحفاظ عليها. • منح المرؤوسين السلطات والمسؤوليات الكافية لأداء عملهم. • التواصل بشكل واضح ومفهوم عند اصدار التعليمات والتوجيهات. • القدرة على التعامل مع المراجعين بلطف واحترام وبشكل واضح ومفهوم.
مهام الوظيفة (Service Functions)	
مهام الوظيفة Service Functions	• الالتزام بتنفيذ جميع الأعمال التي من شأنها تسهيل عمل المؤسسة في سبيل تحقيق أهدافها وتطوير كل ما من شأنه تحسين العمل.
أهداف الوظيفة (Objectives)	
الأهداف Objectives	• الإيمان بأهداف المؤسسة والعمل بتفان لتحقيق هذه الأهداف وتنفي الخطط والتعليمات الموضوعة في سبيل ذلك والالتزام بتنفيذ القوانين والمراسيم والقرارات والأنظمة المتعلقة بعمل المؤسسة.

3.1.2. تجهيز معطيات البحث (A Research Data Preparation)

- كما ذكر سابقاً فقد تم بناء مستودع بيانات يحوي بيانات لعاملين من مؤسسات ومنظمات مختلفة. يتألف مستودع البيانات من بيانات /100/ عامل من مختلف الفئات الوظيفية تشكل البيانات الأصلية لهذا البحث (Original Data) والمؤلفة من:

1. المعلومات الأساسية كالاسم والفئة الوظيفية والشهادة والعمر والجنس وعدد سنوات الخدمة.

2. نتائج تقييم الأداء نصف السنوي لعام 2013 بعد نقلها من نماذج التقييم الخاصة بالمؤسسة التي ينتمي إليها الموظف إلى نموذج التقييم الموحد, وبالتالي فإن نتائج التقييم المعتمدة في هذا البحث هي نتائج التقييم الموجودة ضمن النموذج الموحد وعدد /200/ سجل.
3. بيانات التفقد اليومي لعام 2013 والمؤلفة من بيانات متناثرة مخزنة ضمن قواعد بيانات وسجلات مختلفة غير مترابطة وبالتالي غير صالحة للاستخدام داخل بيئة عمل التنقيب في المعطيات. بغية إيجاد هيكلية مناسبة للاستفادة من هذه البيانات القيمة فقد تم تجميعها حسب الشكل التالي:

- ✓ عدد أيام الغياب (Count of Absence Days): تتضمن عدد أيام الغياب غير المشروع وعدد أيام الاجازات الخاصة بدون راتب أو الاستيداع خلال عام 2013.
- ✓ نسبة الاستفادة من الاجازات الإدارية المسموح بها (Leave Utilization Rate): تتضمن نسبة الاستفادة من عدد أيام الإجازات الإدارية المسموح بها خلال عام 2013.
- ✓ عدد أيام الاستراحات المرضية (Count of Rest Days): تتضمن عدد أيام النقاهات والاستراحات المرضية خلال عام 2013.

- إن نتائج التقييم حسب النموذج الموحد غير صالحة بعد للاستخدام كعينة تدريبية وهي بحاجة لإجراء عدد من عمليات التصفية للحصول على مجموعة بيانات متناسقة تدعى بيانات التدريب (Training Data), وقد شملت عمليات التصفية التي تم اجراؤها العمليات التالية:

1. استبعاد جميع عاملي الفئة الخامسة وعاملي الجهد العضلي من عمال الفئة الرابعة والذين يتم تصنيفهم ضمن الوظائف الإنتاجية وعمليات الإنتاج المباشر نظراً لأن هذه الوظائف لم يتم تشميلها ضمن هذا البحث.
2. محاولة التقريب والتوحيد بين نقاط التقييم بغية التقليل من تناثر البيانات وكمثال على ذلك يمكن أن نلاحظ أن القيمة الدنيا لبيانات الوظيفة (F) هي(2) ولا وجود لبيانات تتراوح قيمتها بين (0-1.9)

وذلك بغية تجمع البيانات ضمن نطاق معين وجعل فرزها ذو دقة أعلى نظراً لقلة عدد البيانات المتاحة نسبياً.

3. استبعاد جميع البيانات المتناقضة, وكمثال على ذلك بيانات التقييم لأشخاص ضمن الوظيفة (A) ونقاط تقييم أدائهم أقل من نقاط تقيم أداء أشخاص ضمن الوظيفة (B) أو الوظيفة (F).
4. تعديل جميع البيانات المتشابهة بشكل كامل كونها غير ذات فائدة في التصنيف وذلك من خلال تعديل بعض القيم بنسب بسيطة غير ذات تأثير ضمن مستودع البيانات الحالي ولكن مستقبلاً ومع توسع هذه المستودع تصبح ذات فائدة وتساهم في عمليات التصنيف.

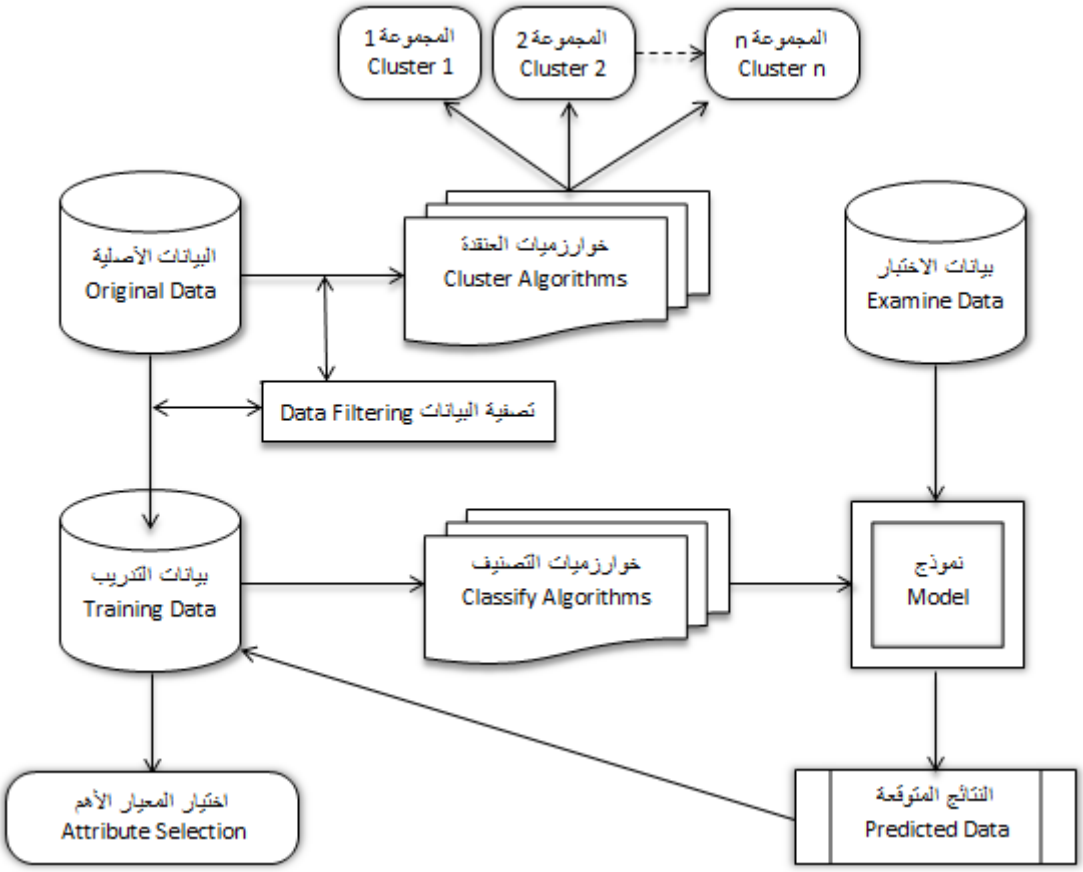
- بعد عمليات التصفية تم استخلاص بيانات التدريب وعددها /100/ تقييم نصف سنوي مختلف مبينة في الملحق رقم (6), تم توزيع بيانات حسب الوظائف وفق ما يلي:
  - /39/ عامل من الوظيفة A.
  - /44/ عامل من الوظيفة B.
  - /17/ عامل من الوظيفة F.
- تمت الاستفادة من بعض نتائج التقييم التي تم استبعادها في عملية التصفية من أجل بناء معطيات اختبار (Examine Data) لاختبار النموذج التنبؤي في التكهن بالوظيفة المناسبة لهذه المعطيات, تتألف معطيات الاختبار من قسمين:
  - بيانات ذات قيم منتظمة (Regular Value): وهي بيانات قريبة نسبياً من بيانات التدريب (Training Data) ولا تحتوي بيانات ذات قيم شاذة كما أنها ضمن مجال (Range) متناسب مع مجال بيانات التدريب.
  - بيانات ذات قيم غير منتظمة (Irregular Value): وهي بيانات ذات تشويش (Noise) عالي وبعيدة عن مجال بيانات التدريب.

2.2. التنقيب في المعطيات (Data Mining)

للاستفادة من المعطيات التي تم جمعها ضمن بيئة التنقيب في المعطيات من خلال بناء النماذج التنبؤية بالاعتماد على خوارزميات التصنيف ومن ثم انتقاء أهم معايير تقييم الأداء استناداً لخوارزميات انتقاء المعايير وأخيراً تحليل بيانات عملية التفقد اليومي بالاستفادة من خوارزميات العنقدة.

1.2.2 بناء نموذج العمل (Build a Working Model)

- تهدف هذه الخطوة لبناء نموذج العمل التنبؤي الموضح في الشكل (2–1) بالاستفادة من بيانات التدريب (Training Data) لبناء نماذج تنبؤية (Predictive Models) من خلال استخدام تقنيات التصنيف التي تم توضيحها في القسم النظري, ومن ثم الاستفادة بشكل عملي من هذه النماذج في التنبؤ بالوظيفة المناسبة لعدد من بيانات الاختبار (Examine Data) التي تم توضيحها سابقاً.
- لكل سجل من بيانات الاختبار يوجد خمس نتائج تم التكهن بها وذلك بسبب استخدام خمس تقنيات مختلفة للتصنيف, وبالتالي باعتماد مبدأ التصويت للأكثرية (Majority Voting Scheme) يمكن تحديد الوظيفة المناسبة وفي حال إجماع النتائج الخمس على قيمة واحدة سيتم إضافتها إلى بيانات التدريب مما يسمح بالتعليم الذاتي (Self-Learning) لنموذج العمل.
- تتم الاستفادة من نتائج التصنيف في اختيار أهم المعايير التي اعتمدتها التصنيفات المختلفة في بناء النماذج (Models) والتي تعبر عن أهم معايير تقييم أداء العاملين والتي يجب أخذها بعين الاعتبار عند انتقاء الموظف او في عملية التوظيف واستخدام العاملين.
- أما بالنسبة لمعطيات التفقد اليومي والتي تم دمجها مع بعض لمعلومات الأساسية للعاملين فسيتم معالجتها عبر خوارزميات العنقدة لبناء مجموعات ذات معنى ومعلومات مفيدة عن هذه البيانات.



الشكل 1.2 : النموذج المقترح لمعالجة المعطيات التي تم جمعها /البيانات الأصلية/.

- تم استخدام تطبيق WEKA لتنفيذ وبناء نموذج العمل الموضح في الشكل (1.2) وفق منهجين هما:

1) المنهج رقم /1/: اعتماد كافة المعايير وعددها /21/ معياراً تتضمن استخدام كافة معايير تقييم الأداء في النموذج الموحد أي معايير الكفاءات الأساسية (Basic Skills) وعددها /18/ معياراً بالإضافة إلى معياري مهام الوظيفة (Service Function) ومعيار أهداف الوظيفة (Objective) ومعيار نوع الوظيفة (A, B, F).

2) المنهج رقم /2/: حساب معدل عوامل الكفاءات الرئيسية والمؤلف من /18/ عامل مختلف وفق المعادلة:

$$Basic - Skills = \frac{1}{18}\sum_{k=0}^{17} \text{attribute(k)}$$

واعتماد أربعة معايير هي: الكفاءات الأساسية (Basic-Skills) ومهام الوظيفة (Service Functions) وأهداف الوظيفة (Objectives) بالإضافة إلى معيار نوع الوظيفة (A, B, F).

- من أجل تحديد دقة تصنيف البيانات التي تم تدريبها تم اعتماد أسلوب الاختبار (Cross-validation) بنسبة تقطيع مساوية /10/ (Folds = 10) وذلك لجميع حالات التصنيف في هذا البحث ويمكن تلخيص عمل هذه الطريقة بالخطوات التالية:

1) تدريب نسبة 100% من بيانات التدريب.
2) تقسيم بيانات التدريب إلى /10/ أقسام بشكل عشوائي.
3) تنفيذ عملية تخمين لمقادير لكل من هذه الأقسام العشرة وحساب نسبة تطابق القيم الحقيقية لهذه البيانات مع القيم التي تم تخمينها.
4) حساب هذه النسبة لكامل مجموعة البيانات لتكون نسبة الدقة المطلوبة

- يجب التنويه هنا أن مقدار نسبة التقسيم مرتبط بعدد بيانات التدريب ولا يوجد نسبة ثابتة تحدد الخيار الأفضل للتقسيم ولكن الافتراضي المستخدم ضمن تطبيق WEKA هو /10/ حيث أشارت العديد من الدراسات إلى أن هذه النسبة شكلت القيمة الأمثل في العديد من الاختبارات.

2.2.2. نتائج المنهج رقم /1/

- تم جدولة البيانات المرتبطة بالمنهج رقم /1/ ضمن الملف ذو اللاحقة (arff) المبين في الملحق رقم(6) المتضمن بيانات التدريب وعددها /100/ سجل موزعة على /21/ معيار.
- تم اجراء خوارزميات التصنيف التي تم اعتمادها في هذا البحث وجدولة النتائج في الجدول (2.2).
- كما يبين الشكل رقم (2.2) جدول مصفوفات الشك (Confusion Matrix) الناتجة عن تطبيق المنهج رقم /1/ حسب خوارزميات التصنيف المعتمدة في هذا البحث.

جدول 2.2 : نتائج خوارزميات التصنيف حسب المنهج رقم /1/

	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Time (s)
Naïve Bayes	94%	6%	0.9036	0.02
J48	90%	10%	0.8376	0.02
MLP	98%	2%	0.9679%	0.97
KNN	86%	14%	0.7736%	0
SVM	93%	3%	0.8876	0.13

J48

التصنيف الحقيقي \ التصنيف المتوقع	A	B	C
A	37	2	0
B	4	40	0
C	0	4	13

Naïve Bayes

التصنيف الحقيقي \ التصنيف المتوقع	A	B	C
A	36	3	0
B	2	42	0
C	0	1	16

MLP

التصنيف الحقيقي \ التصنيف المتوقع	A	B	C
A	39	0	0
B	1	43	0
C	0	1	16

KNN

التصنيف الحقيقي \ التصنيف المتوقع	A	B	C
A	35	4	0
B	5	38	0
C	1	3	13

SVM

التصنيف الحقيقي \ التصنيف المتوقع	A	B	C
A	36	3	0
B	3	41	0
C	0	1	16

الشكل 2.2 : مصفوفات الشك الناتجة عن تطبيق المنهج رقم /1/ حسب خوارزميات التصنيف المعتمدة.

- باختيار عينة من بيانات الاختبار والمبينة في الملحق رقم (7), تم تنفيذ عملية التنبؤ لهذه البينات حسب النماذج الخمس المعتمدة في هذا البحث ومن ثم جدولة النتائج في الجدول (3.2).

الجدول 3.2 : النتائج المتوقعة لبيانات الاختبار حسب المنهج رقم /1/

	Regular Testing Data						Irregular Testing Data					
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Naïve Bayes	A	B	F	F	B	B	B	F	F	B	F	F
J48	A	B	F	F	B	B	B	F	B	B	B	F
MultiLayerPerceptron	A	B	F	F	B	B	A	F	B	B	B	B
KNN	A	B	F	F	B	B	B	B	B	B	A	B
SVM	A	B	F	F	B	B	B	F	B	B	B	F

2.2.2. نتائج المنهج رقم /2/

- حساب معدل عوامل الكفاءات الرئيسية والمؤلف من /18/ عامل مختلف وفق المعادلة:

$$Basic - Skills = \frac{1}{18}\sum_{k=0}^{17} \text{attribute(k)}$$

- اعتماد العوامل الثلاث: المهارات الرئيسية (Basic-Skills) ومهام الوظيفة (Service-Functions) وأهداف الوظيفة (Organization-Objectives), بالإضافة إلى متغير الوظيفة (Result) الذي يحتوي نوع الوظيفة التي يشغلها الموظف (A, B, F).
- تم جدولة البيانات المرتبطة بالمنهج رقم /2/ ضمن الملف الموضح في الملحق (8) المتضمن بيانات التدريب وعددها /100/ سجل موزعة على /4/ معايير.
- تم اجراء خوارزميات التصنيف التي تم اعتمادها في هذا البحث وجدولة النتائج في الجدول (4.2).

جدول 4.2 : نتائج خوارزميات التصنيف حسب المنهج رقم /2/

	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Time (s)
Naïve Bayes	96%	4%	0.936	0
J48	94%	4%	0.9042	0.01
MLP	95%	5%	0.9201	0.16
KNN	97%	3%	0.9518	0
SVM	97%	3%	0.9518	0.01

- كما يبين الشكل (3.2) جداول مصفوفة الشك الناتجة عن تطبيق المنهج رقم /2/ حسب خوارزميات التصنيف المعتمدة في هذا البحث.

Naïve Bayes		التصنيف المتوقع		
		A	B	C
التصنيف الحقيقي	*A*	38	1	0
	B	1	42	1
	C	0	1	16

J48		التصنيف المتوقع		
		A	B	C
التصنيف الحقيقي	*A*	38	1	0
	B	3	40	1
	C	0	1	16

MLP		التصنيف المتوقع		
		A	B	C
التصنيف الحقيقي	*A*	38	1	0
	B	2	41	1
	C	0	1	16

KNN		التصنيف المتوقع		
		A	B	C
التصنيف الحقيقي	*A*	38	1	0
	B	1	43	0
	C	0	1	16

SVM		التصنيف المتوقع		
		A	B	C
التصنيف الحقيقي	*A*	38	1	0
	B	1	43	0
	C	0	1	16

الشكل 3.2 : مصفوفات الشك الناتجة عن تطبيق المنهج رقم /2/ حسب خوارزميات التصنيف المعتمدة.

- باختيار عينة من بيانات الاختبار والمبينة في الملحق رقم (9), تم تنفيذ عملية التنبؤ لهذه البينات حسب النماذج الخمس المعتمدة في هذا البحث ومن ثم جدولة النتائج في الجدول (4.2).

الجدول 4.2 : النتائج المتوقعة لبيانات الاختبار حسب المنهج رقم /2/

	Regular Testing Data						Irregular Testing Data					
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Naïve Bayes	A	B	F	B	B	B	F	B	A	B	F	B
J48	A	B	F	B	B	B	F	B	A	F	B	F
MLP	A	B	B	B	B	B	B	A	B	F	B	F
KNN	A	B	F	B	B	B	A	B	A	B	B	B
SVM	A	B	F	B	B	B	F	A	B	B	B	B

3.2. انتقاء المعايير (Attribute Selection)

- بالاعتماد على الخوارزميات المتعددة التي تسمح باختيار المعايير ذات الأهمية في عمليات التصنيف (Attribute Selection) تم إجراء عدد من عمليات اختيار أهم معايير التقييم وتم ترتيب النتائج في الجدول(6.2).

الجدول (6.2): نتائج خوارزميات انتقاء المعايير نتائج عملية تقييم أداء العاملين

مقيم المعيار Attribute Evaluator	منهج البحث Search Method	نتيجة اختيار المعيار Attribute Selection Output	
OneRAttributeEval	Ranker	Org. Objectives	94
		Service Functions	93
		Execution	85
		Quality	82
		Interaction-with-Coworkers	81
		Interaction-with-E/C	80
		Self-Development	79
		Safety	78
		Knowledge	78
		Economic	78
		Interaction-with-Supervisors	77
		Productivity	77
		Problem-Solving-and-Resolution	76
		Organization	74
		Politeness	74
		Initiative-and-Innovation	73
		Strategic-Thinking	71
		Attendance-and-Punctuality	71
		Adaptability	71
		Personal-Appearance	55

CfsSubsetEval	GeneticSearch	Execution Safety Productivity Initiative-and-Innovation Strategic-Thinking Interaction-with-E/C Service Functions Org. Objectives
FilteredSubsetEval	GreedyStepwise	Service Functions Org. Objectives

4.2. معالجة بيانات التفقد اليومي (Build a Working Model)

- كم هو موضح في الشكل (1.2) فإن البيانات الأصلية (Original Data) تحوي معطيات التفقد اليومي الخاص بالعاملين التي سيتم معالجتها من خلال خوارزميات التجزئة في تطبيق WEKA, حيث تعمل خوارزميات التجزئة المختلفة على تجميع المعطيات ضمن مجموعات بناء على نسبة تشابهها وتقاربها ويتم تحديد عدد المجموعات بناء على مدى المعلومات المفيدة التي يمكن استخراجها من هذه المجموعات.
- بناء على المعطيات المتعلقة بالتفقد اليومي للعاملين والتي تم توضيحها سابقاً فقد تم تقسيم وجدولة هذه المعطيات وإدراجها ضمن المعايير التالية:
  - الفئة الوظيفية (Level) التي يندرج ضمنها الموظف.
  - الجنس (Gender) سواء ذكر أو أنثى.
  - الدرجة العلمية (Certificate) للعامل.
  - عدد سنوات الخدمة (Service Duration).
  - عدد أيام الغياب غير المبرر (Count of Days Absence).
  - نسبة الاستفادة من أيام الإجازات الإدارية (Leave Utilization Rate).
  - عدد أيام الاستراحات المرضية (Count of Rest Days).

- تم اجراء خوارزمية العنقدة (Simple K-Means) مع قيم متعددة للمتغير (K) لتغيير عدد المجموعات المراد توزيع بيانات التفقد اليومي ضمنها, يظهر الجدول (7.2) ناتج تجزئة البيانات إلى ثلاث مجموعات, بينما يبين الجدول (8.2) ناتج تجزئة البيانات إلى أربع مجموعات أما الجدول (9.2) فهو يبين توزع نتائج التجزئة إلى خمس مجموعات.

الجدول 7.2 : توزع بيانات التفقد ضمن ثلاث مجموعات (k=3).

Attribute	Clusters		
	0	1	2
	46	41	13
Gender	M	M	F
Certificate	Bachelor	Institute	Bachelor
Service Duration	23.6304	10.7561	18.6154
Count of Days Absence	1.413	2.8049	2.3846
Leave Utilization Rate	41.5217	88.0488	71.5385
Count of Rest Days	8.0435	12.0732	20.7692

الجدول 8.2: توزع بيانات التفقد ضمن أربع مجموعات (k=4).

	Clusters			
Attribute	0	1	2	3
	37	29	17	17
Gender	M	M	M	F
Certificate	Bachelor	High School	Bachelor	Bachelor
Service Duration	24.7568	11.6207	13.7647	16.6471
Count of Days Absence	1.2432	1.7931	4.2353	2.4118
Leave Utilization Rate	34.0541	79.6552	92.9412	76.4706
Count of Rest Days	7.4324	8.1034	17.3529	19.4118

الجدول 9.2 : توزع بيانات التفقد ضمن خمس مجموعات (k=5).

	Clusters				
Attribute	0	1	2	3	4
	30	21	15	17	17
Gender	M	M	M	F	M
Certificate	Bachelor	High School	Bachelor	Bachelor	Educated
Service Duration	25.8333	10.6667	13.3333	16.6471	16.9412
Count of Days Absence	1.1	1.8095	3.8667	2.4118	2.4118
Leave Utilization Rate	33.6667	78.0952	92.6667	76.4706	65.2941
Count of Rest Days	7.6667	8.3333	15	19.4118	10.2941

الفصـــــــل الثـــالث

المناقشـــــة

تمهيد

يهدف هذا الفصل لتحليل النتائج التي تم الحصول عليها في الفصل السابق لاكتشاف نقاط القوة ونقاط الضعف للنموذج المقترح وتحديد الشروط المثلى لاستخدام التنقيب في معطيات فعاليات الكوادر البشرية.

1.3. نتائج المنهج رقم /1/

إن قوة أداء المصنف تنبع من الدقة في تصنيف بيانات التدريب وبالتالي ارتفاع دقة التوقع, ومع اعتماد طريقة (Cross-Validation) بنسبة تقطيع /Fold=10/ لجميع البيانات التي تم تدريبها فإنه يمكن اعتبار نسبة عدد حالات التصنيف بشكل صحيح (Correctly Classified Instances) ونسبة عدد حالات التصنيف غير الصحيح (Incorrectly Classified Instances) بالإضافة إلى قيمة المتغير الإحصائي كابا (Kappa) هما المؤشر الأول لأداء خوارزمية التصنيف بينما يرتبط المؤشر الثاني لأداء خوارزمية التصنيف بالتوقع الفعلي لبيانات الاختبار.

1.1.3. دقة التصنيف بالنسبة لبيانات التدريب

بإعادة ترتيب نتائج هذا المنهج ترتيباً تنازلياً ضمن الجدول (1.3) نلاحظ أن دقة خوارزمية (MLP) هي الأعلى بين خوارزميات التصنيف الأخرى مع الإشارة إلى أنها استغرقت الوقت الأطول لبناء النموذج الخاص بها[41]. كما يمكن ملاحظة ضعف أداء خوارزمية الجار الأقرب (KNN) واحتلالها المرتبة الأخيرة بينما كان أداء الخوارزميات المتبقية (Naïve Bayes, SVM, J48) متقارباً.

الجدول 1.3 : ترتيب نتائج خوارزميات التصنيف للمنهج /1/ حسب دقة التصنيف.

	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Time (s)
MLP	98%	%2	0.9679%	0.97
Naïve Bayes	94%	4%	0.9036%	0.02
SVM	93%	3%	0.8876%	0.13
J48	%90	%10	0.8376%	0.02
KNN	%86	%14	0.7736%	0

[41] غالباً ما تستغرق هذه الخوارزمية زمناً أطول من باقي الخوارزميات نظراً لآلية عمل هذه الخوارزمية التي تم توضيحها في قسم التمهيد النظري والتي تحاول إيجاد الحالة الأمثل.

للتعمق أكثر في تباين أداء خوارزميات التصنيف الخمس يجب العودة إلى مصفوفات الشك الخاصة بها والمبينة في الشكل (1.3), حيث يلاحظ أن مصفوفة الشك الخاصة بخوارزمية (MLP) قامت بالتصنيف بالشكل الأفضل وكان الخطأ بتوقعين فقط كما هو مبين في الشكل (1.3.أ), أما بالنسبة للخوارزمية (KNN) ذات الدقة الأدنى فنلاحظ من الشكل (1.3.هـ) أن مقدار الخطأ في التوقع كبير جداً حيث يوجد توقعات خاطئة عدد /6/ من أصل /44/ خاصة بالوظيفة /B/, وتوقعات خاطئة عدد /4/ خاصة بالوظيفة /A/ وتوقعات خاطئة عدد /1/ خاصة بالوظيفة /F/ كما تظهر حالة شاذة لخوارزمية (KNN) مبينة في الشكل (1.3.هـ) باللون الأحمر وتبين مدى الخطأ الكبير في التوقع, حيث تم التوقع لأحد قيم الوظيفة F بقيمة الوظيفة A بدلاً منها.

يمكن تفسير النسبة العالية للتوقعات الخاطئة لهذا التصنيف بآلية عمل هذه الخوارزمية التي تعمد لحساب احتمال كل من الأحداث بالاعتماد على /k/ حدث مجاور, وبالتالي انخفاض عدد بيانات التدريب بالنسبة لعدد المعايير ضمن هذا المنهج جعل هذه الخوارزمية غير قادرة على بناء النموذج المناسب وبالتالي ارتفاع نسبة التوقعات الخاطئة.

MLP		التصنيف المتوقع		
		A	B	F
التصنيف الحقيقي	*A*	39	0	0
	B	1	43	0
	F	0	1	16

(أ)

Naïve Bayes		التصنيف المتوقع		
		A	B	F
التصنيف الحقيقي	*A*	36	3	0
	B	2	42	0
	F	0	1	16

(ب)

J48		التصنيف المتوقع		
		A	B	F
التصنيف الحقيقي	*A*	37	2	0
	B	4	40	0
	F	0	4	13

(ج)

SVM		التصنيف المتوقع		
		A	B	F
التصنيف الحقيقي	*A*	36	3	0
	B	3	41	0
	F	0	1	16

(د)

KNN		التصنيف المتوقع		
		A	B	F
التصنيف الحقيقي	*A*	35	4	0
	B	6	38	0
	F	**1**	3	13

(هـ)

الشكل 1.3: مصفوفات الشك للمنهج رقم /1/, مرتبة تنازلياً

أما بالنسبة لخوارزميات التصنيف (Naïve Bayes) و (SVM) و (J48) فقد كان أداؤها متقارباً من حيث الدقة وجميع الأخطاء في التصنيف كانت بين وظيفتين متجاورتين أي (A بدلاً B).

2.1.3. دقة التوقع بالنسبة لبيانات الاختبار

بإعادة ترتيب توقعات بيانات الاختبار ترتيباً تنازلياً حسب دقة التوقع كما هو مبين في الجدول (2.3) يمكن ملاحظة تطابق نتائج التوقع بشكل كامل بين جميع خوارزميات التصنيف عند استخدام عينة بيانات الاختبار المنتظمة (Regular Examine Data). بينما في حال استخدام عينة بيانات الاختبار غير المنتظمة (Irregular Examine Data) يلاحظ وجود تطابق في النتائج المتوقعة لبعضها كالمجموعة (M10) وتفاوت في بعضها الآخر, هذا التفاوت ضئيل في بعض المجموعات مثل (M7, M8, M9), بينما يزداد التفاوت في المجموعات الأخرى وخاصة مجموعة البيانات /M11/ حيث تراوحت التخمينات بين الفئات الثلاثة (A, B, F) وهذا ما يشكل تعارض غير منطقي. بالتدقيق في الجدول (3-2) يلاحظ أن التوقع ذو القيمة الشاذة مرتبط بخوارزمية (KNN) ذات الدقة الأدنى التي تم الحصول عليها[42] لدى اتباع هذا المنهج كما هو موضح الفصل السابق وبالتالي فإن الاختبار العملي لأداء المصنف KNN أثبت عدم قدرته على بناء نموذج تنبؤي جيد بالنسبة لهذا النوع من المعطيات.

[42] حسب Cross-Validation

الجدول 2.3 : النتائج المتوقعة لبيانات الاختبار حسب مبدأ تصويت الأكثرية.

	Regular Examine Data						Irregular Examine Data					
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
MLP	A	B	F	F	B	B	A	F	B	B	B	B
Naïve Bayes	A	B	F	F	B	B	B	F	F	B	F	F
SVM	A	B	F	F	B	B	B	F	B	B	B	F
J48	A	B	F	F	B	B	B	F	B	B	B	F
KNN	A	B	F	F	B	B	B	B	B	B	A	B
Majority Result	**A**	**B**	**F**	**F**	**B**	**B**	**B**	**F**	**B**	**B**	**B**	**F**

باعتماد مبدأ التصويت للأكثرية يمكن اختيار الوظيفة المناسبة أو التوقع المناسب لعينة الاختبار المطلوبة, وقد تم ادراج القيمة المتوقعة في الجدول (2.3) كنتيجة لهذه العملية (Majority Result).

2.3. نتائج المنهج رقم /2/

كما ذكر في مناقشة نتائج المنهج رقم /1/ سيتم اعتبار نسبة عدد حالات التصنيف بشكل صحيح ونسبة عدد حالات التصنيف غير الصحيح بالإضافة إلى قيمة المتغير الإحصائي كابا (Kappa) هما المؤشر الأول لأداء خوارزمية التصنيف بينما يرتبط المؤشر الثاني لأداء خوارزمية التصنيف بالتوقع الفعلي لبيانات الاختبار حسب مبدأ التصويت للأغلبية.

1.2.3. دقة التصنيف (بيانات التدريب)

بإعادة ترتيب نتائج هذا المنهج ترتيباً تنازلياً ضمن الجدول (3.3) يبدو أن دقة خوارزمية (SVM) هي الأعلى تليها خوارزمية (KNN) ثم خوارزمية (Naïve Bayes), ثم خوارزمية (MLP), وفي النهاية خوارزمية (J48). كما استغرقت خوارزمية (MLP) زمناً أطول من باقي الخوارزميات لبناء النموذج الخاص بها.

الجدول 3.3 : ترتيب نتائج خوارزميات تصنيف المنهج /2/ حسب دقة التصنيف.

	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Time (s)
SVM	97 %	3 %	0.9518	0.01
KNN	97 %	3 %	0.9518	0
Naïve Bayes	96%	4%	0.936	0
MLP	95 %	5 %	0.9201	0.16
J48	94 %	4 %	0.9042	0.01

بالأخذ بعين الاعتبار آلية عمل خوارزمية (SVM) التي تعمد إلى إيجاد أفضل سطح فاصل يتوسط جميع النقاط ومع وجود ثلاثة معايير فقط ومائة سجل ذات بيانات متناسقة وقليلة الشذوذ فإن هذه الخوارزمية تعمل بشكل جيد وقد ارتفعت الدقة الخاصة بهذا المصنف ضمن هذا المنهج بمقدار 4%.

إن دقة التصنيف (KNN) ارتفعت من 86% إلى 97% وذلك لوجود عدد أكبر من بيانات التدريب بالنسبة لعدد المعايير مما يمنح هذا المصنف قدرة أكبر على الفصل بين البيانات.

كما يلاحظ تحقيق خوارزمية (Naïve Bayes) وخوارزمية (MLP) نسبة دقة متقاربة هي 96% و 95%, بينما دقة خوارزمية (J48) هي الأدنى وبلغت 94%.

بالعودة إلى مصفوفات الشك الخاصة بخوارزميات التصنيف والمبينة في الشكل (2.3) يلاحظ أن التوقع الخاطئ في هذا المنهج متعادل بشكل أفضل من التوقع الخاطئ لدى المنهج رقم /1/ وذلك لتناسب عدد بيانات التدريب مع عدد المعايير. كما تتطابق مصفوفتي الشك بين المصنفات (SVM و KNN) الأعلى دقة كما يظهر في الشكل رقم (2.3.أ) والشكل رقم (2.3.ب), بينما نلاحظ ارتفاع نسبة التوقع الخاطئ بين الوظيفة /A/ والوظيفة /B/ كما يظهر في الشكل رقم (2.3.د) والشكل (2.3.هـ).

SVM		التصنيف المتوقع		
		A	B	C
التصنيف الحقيقي	*A*	38	1	0
	B	1	43	0
	C	0	1	16

(أ)

KNN		التصنيف المتوقع		
		A	B	C
التصنيف الحقيقي	*A*	38	1	0
	B	1	43	0
	C	0	1	16

(ب)

Naïve Bayes		التصنيف المتوقع		
		A	B	C
التصنيف الحقيقي	*A*	38	1	0
	B	1	42	1
	C	0	1	16

(ج)

MLP		التصنيف المتوقع		
		A	B	C
التصنيف الحقيقي	*A*	38	1	0
	B	2	41	1
	C	0	1	16

(د)

J48		التصنيف المتوقع		
		A	B	C
التصنيف الحقيقي	*A*	38	1	0
	B	3	40	1
	C	0	1	16

(هـ)

الشكل 2.3 : مصفوفات الشك لنتائج تصنيف المنهج /2/, مرتبة ترتيباً تنازلياً.

2.2.3. دقة التوقع بالنسبة لبيانات الاختبار

بإعادة ترتيب توقعات بيانات الاختبار تنازلياً حسب دقة التصنيف كما هو مبين في الجدول (4.3) يمكن ملاحظة تطابق نتائج التوقع بشكل كامل بين جميع خوارزميات التصنيف عند استخدام عينة بيانات الاختبار المنتظمة. بينما في حال استخدام عينة بيانات الاختبار غير المنتظمة يلاحظ عدم وجود أي تطابق في النتائج المتوقعة وإنما تفاوتت فيما بينها, هذا التفاوت ضئيل في بعض المجموعات مثل (M8, M9), بينما يزداد التفاوت في المجموعات الأخرى وخاصة مجموعة البيانات (M7) حيث تراوحت التخمينات بين الفئات الثلاثة (A, B, F) وهذا ما يشكل تعارض غير منطقي كالذي سبق وحدث لدى تحليل نتائج المنهج رقم /1/

وبالتدقيق في الجدول (4.3) يلاحظ أن التنبؤ الشاذ مرتبط بخوارزمية (KNN) ذات الدقة الجيدة في هذا المنهج بالنسبة لقياس الدقة حسب (Cross-Validation).

الجدول 4.3 : النتائج المتوقعة لبيانات الاختبار حسب مبدأ تصويت الأكثرية

	Regular Testing Data						Irregular Testing Data					
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
SVM	A	B	F	B	B	B	F	A	B	B	B	B
KNN	A	B	F	B	B	B	***A***	B	A	B	B	B
Naïve Bayes	A	B	F	B	B	B	F	B	A	B	F	B
MLP	A	B	B	B	B	B	B	A	B	F	B	F
J48	A	B	F	B	B	B	F	B	A	F	B	F
Majority Result	**A**	**B**	**F**	**B**	**B**	**B**	**F**	**B**	**A**	**B**	**B**	**B**

3.3. مقارنة نتائج المنهجين /1/ و/2/:

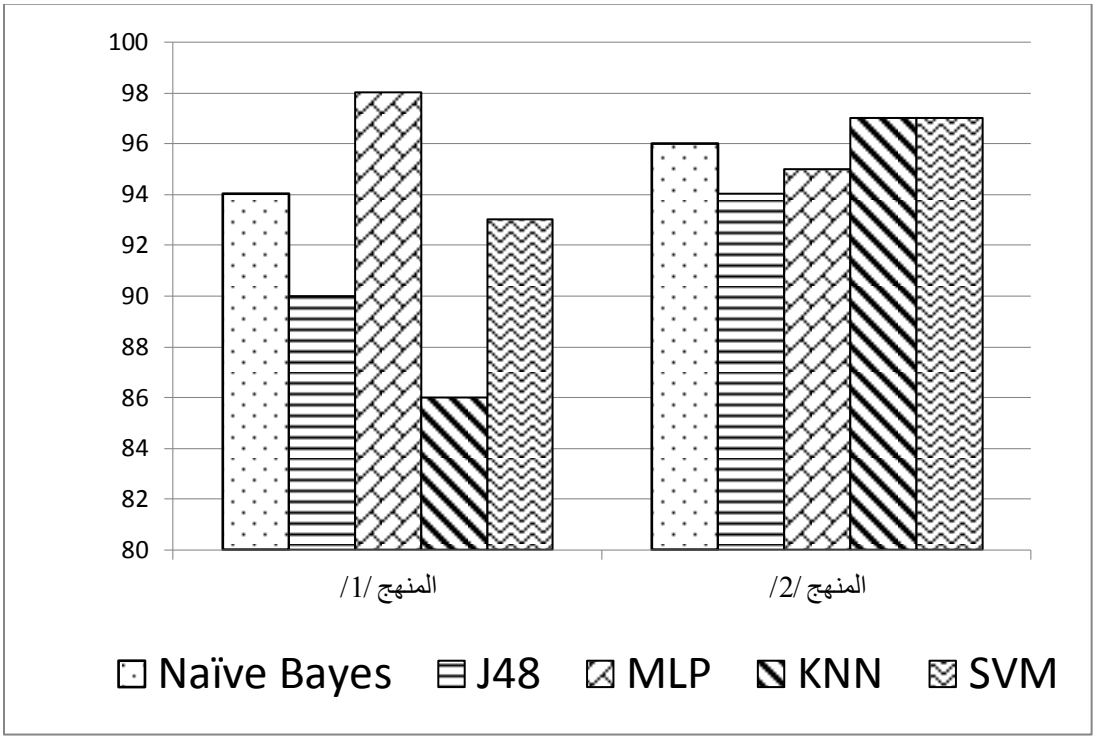
للمقارنة بين نتائج المنهجين لابد من التذكير بأن عدد بيانات التدريب للمنهجين هو /100/ سجل بينما تفاوت عدد معايير التقييم حسب الشكل التالي:

1. المنهج /1/ يتألف من /21/ معيار للتقييم تشمل معايير الكفاءات الاساسية وعددها /18/ معيار بالإضافة إلى معيار مهام الوظيفة ومعيار أهداف الوظيفة ومعيار الفئة الوظيفية.
2. المنهج رقم /2/ يتألف من /4/ معايير تشمل الكفاءات الأساسية والتي تم حسابها كمعدل عن نقاط تقييم الكفاءات الأساسية, بالإضافة إلى مهام الوظيفة وأهداف الوظيفة ومعيار الفئة الوظيفية.

من خلال التحليل السابق لنتائج المنهجين ومقارنة مصفوفات الشك والمعامل الاحصائي ونتائج التنبؤ لبيانات الاختبار وبالتوجه للمخطط البياني (3.3) المتضمن مقارنة نتائج المصنفات يمكن ملاحظة ما يلي:

1. نسبة تفاوت دقة نماذج المنهج /2/ أقل منها في المنهج /1/, حيث يلاحظ أن نتائج المنهج /1/ تراوحت ما بين 86% إلى 98% بينما نتائج المنهج /2/ تراوحت ما بين 94% إلى 97% ويمكن تفسير ذلك بأن عدد البيانات أصبح أكثر تناسباً مع عدد المعايير.

2. ارتفاع دقة جميع المصنفات عند تطبيق المنهج /2/ ما عدا نموذج المصنف (MLP), حيث يلاحظ انخفاض الدقة من 98% إلى 95% ويعلل ذلك بأهمية عدد المعايير بالنسبة لهذا المصنف بشكل أكبر من أهميتها بالنسبة لباقي المصنفات[43] وذلك لطبيعة عمل هذا المصنف المستند إلى خوارزمية الشبكات العصبونية حيث يساهم العدد الأكبر من المعايير ببناء شبكة عصبونية أوسع تكون أكثر دقة في تصنيف بيانات التدريب.



الشكل 3.3 : مقارنة نتائج دقة تصنيف المنهج /1/ مع دقة تصنيف المنهج /2/.

3. لا يوجد تفاوت كبير في دقة المصنف (Naïve Bayes) بين المنهجين /1/ و/2/, حيث تراوحت بين 94% و96% ويعلل ذلك بآلية عمل هذه الخوارزمية القائمة على اعتبار المعايير وحدات مستقلة (مبدأ استقلالية الاحداث) وحساب احتمالية كل من هذه المعايير بشكل مستقل عن الآخر وبالتالي فإن عدد المعايير ليس أهمية بالغة عند ثبات عدد بيانات التدريب.

[43] بفرض أن عدد بيانات التدريب ثابت.

4. لا يوجد تفاوت ملحوظ في دقة نموذج المصنف (J48) بين المنهجين /1/ و/2/, حيث تراوحت بين 90% و94% ويعلل ذلك بآلية عمل هذه الخوارزمية المشابهة لآلية عمل خوارزمية (Naïve Bayes) من حيث استقلالية الأحداث مع اعتماد القواعد (Rules) في تحديد القيمة المراد التنبؤ بها.
5. تراوحت دقة الخوارزمية (SVM) بين 93% و97% حيث يمكن وصف أداء هذا المصنف بالمتوازن وهذا متوقع نظراً لقوة أداء هذا المصنف.
6. التفاوت الأكبر في دقة نموذج المصنف (KNN) التي ارتفعت من 86% لدى تطبيق المنهج /1/ إلى 97% عند تطبيق المنهج /2/ وذلك يعود إلى آلية عمل هذه الخوارزمية التي تستند بشكل كامل إلى قيم الأحداث المجاورة وبالتالي فإن عدد المعايير يؤثر بشكل كبير على دقة المصنف[44] وبالتالي فإن إمكانية الاستفادة من هذا المصنف مرتبطة بشرط تناسب عدد المعايير مع عدد بيانات التدريب.
7. لمقارنة نتائج المصنفات بالنسبة للتنبؤ بقيم بيانات الاختبار من الجدولين (2.3) و(4.3) يلاحظ أن نتائج المنهجين /1/ و/2/ متطابقة بين جميع الخوارزميات بالنسبة لبيانات الاختبار المنتظمة, ويمكن وصف أداء خوارزميات التصنيف بالجيد لدى اختبار بيانات متناسبة مع بيانات التدريب وبالتالي يمكن إضافة بيانات الاختبار إلى بيانات التدريب لتوسيع بيانات الاختبار. أما بالنسبة لبيانات الاختبار غير المنتظمة فيلاحظ أن نتائج المصنف (KNN) في كلا المنهجين احتوت نتائج ذات قيم شاذة عن باقي التوقعات الناتجة عن بقية المصنفات وبالتالي فإن ارتفاع دقة هذا المصنف عند تطبيق المنهج /2/ لم تؤثر بشكل كبير على الدقة التنبؤية الضعيفة لهذا المصنف بالنسبة لبينات الاختبار غير المنتظمة والناتجة عن العدد القليل لبيانات التدريب بالنسبة لعدد المعايير. وكنتيجة لتفاوت القيم التي تم التكهن بها يمكن الاعتماد على أسلوب التصويت للأكثرية في ايجاد تقاطع ما بين النتائج التي تم التكهن بها لتحديد الوظيفة الملائمة للعامل ولكن لا يمكن إضافة

[44] بفرض أن عدد بيانات التدريب ثابت.

السجل الخاص به إلى بيانات التدريب لعدم وجود إجماع بين جميع النماذج على نتيجة موحدة. وكنتيجة يمكن القول أن أداء جميع الخوارزميات كان جيداً ما عدا خوارزمية KNN.

4.3. نتائج انتقـــاء المعايير (Select Attribute Results)

بالعودة للجدول (6.2) في الفصل السابق فقد تم استخدام ثلاث خوارزميات مختلفة لانتقاء أهم المعايير التي تم الاستناد إليها في عمليات التصنيف ومن تقاطع نتائج هذه الخوارزميات الثلاث يمكن التأكيد على أن مهام الوظيفة وأهداف الوظيفة تعدان أهم المعايير التي يجب أخذهما بعين الاعتبار في العديد من الفعاليات كالترفيع والترشيح والتوظيف وغيرها, وذلك بناء على خوارزميتي (Genetic و Greedy) بما يشكل تطابق الحالة المثلى العامة مع الحالة المثلى المحلية.

أما بالنسبة للكفاءات الأساسية فيمكن اعتبار (التنفيذ, الأمان, الإنتاجية, المبادرة والإبداع, التفكير الاستراتيجي, التعامل مع الزبائن/المرؤوسين) أهم هذه المعايير ويجب وضعها بالحسبان وإيلاءها أهمية خاصة عند الحاجة للقيام ببعض فعاليات الكوادر البشرية كعمليات الترفيع والترشيح والتوظيف وغيرها.

4.3. نتائج تجزئة بيانات التفقد اليومي

بإعادة ترتيب الجداول (7.2 و 8.2 و 9.2) حسب الشكل (4.3) يمكن ملاحظة آلية عمل خوارزمية (Simple K-mean) خطوة بخطوة مع زيادة عدد المجموعات, هذه المجموعات عبارة عن:

- المجموعة /A/ حافظت بشكل كبير على خصائصها ضمن المراحل الثلاث رغم انخفاض عدد المعطيات المرتبطة بها مع ازدياد عدد المجموعات إلا أنها بقيت المجموعة ذات نسبة المعطيات الأعلى التي يغلب عليها العاملين الذكور من حاملي الشهادات الجامعية وسنوات خدمتهم حوالي /24/ سنة بشكل متوسط. كما أن هناك تقارب كبير في عدد أيام الغياب غير المشروع والتسرب عن العمل

وفي عدد أيام الاستراحات المرضية بينما كان التفاوت أكبر في نسبة الاستفادة من استحقاق الاجازات الإدارية المسموح بها.

Attribute	Cluster#		
Cluster Name	A	B	C
Members Count	46	41	13
Gender	M	M	F
Certificate	Bachelor	Institute	Bachelor
Service Duration	23.6304	10.7561	18.6154
Absence Days Count	1.413	2.8049	2.3846
Leave Utilization Rate	41.5217	88.0488	71.5385
Rest Days Count	8.0435	12.0732	20.7692

Attribute				
Cluster Name	A	E	F	C
Members Count	37	29	17	17
Gender	M	M	M	F
Certificate	Bachelor	High School	Bachelor	Bachelor
Service Duration	24.7568	11.6207	13.7647	16.6471
Absence Days Count	1.2432	1.7931	4.2353	2.4118
Leave Utilization Rate	34.0541	79.6552	92.9412	76.4706
Rest Days Count	7.4324	8.1034	17.3529	19.4118

Attribute					
Cluster Name	A	E	F	C	H
Members Count	30	21	15	17	17
Gender	M	M	M	F	M
Certificate	Bachelor	High School	Bachelor	Bachelor	Educated
Service Duration	25.8333	10.6667	13.3333	16.6471	16.9412
Absence Days Count	1.1	1.8095	3.8667	2.4118	2.4118
Leave Utilization Rate	33.6667	78.0952	92.6667	76.4706	65.2941
Rest Days Count	7.6667	8.3333	15	19.4118	10.2941

الشكل 4.3: توزع البيانات وتشكل المجموعات مع تعديل قيمة عدد المجموعات (K)

- المجموعة /C/ حافظت بشكل كبير على خصائصها ضمن المراحل الثلاث واستمرت كمجموعة ذات نسبة المعطيات الأعلى التي يغلب عليها العاملين الإناث من حاملي الشهادات الجامعية وسنوات خدمتهم حوالي /17/ سنة بشكل متوسط. كما أن هناك تقارب كبير في عدد أيام الغياب غير المشروع

والتسرب عن العمل وفي عدد أيام الاستراحات المرضية بينما كان التفاوت أكبر في نسبة الاستفادة من استحقاق الاجازات الإدارية.

- المجموعة /B/ قد تفرعت إلى مجموعتين تغلب عليهما نسبة العاملين الذكور كما يمكن ملاحظة أن مقدار القيم عند /k=3/ هي الوسطي بين قيم المرتبطة بـ /k=3, k=4/ كما يبين ذلك الجدول رقم (5.3), حيث يظهر أن أغلبية عاملي /k=3/ من حاملي شهادة المعاهد المتوسطة والتي تفرعت عند الانتقال إلى /k=5 , k=4/ إلى حاملي الشهادة الجامعية وحاملي الشهادة الثانوية. وكذلك الأمر بالنسبة لعدد سنوات الخدمة وعدد أيام الغياب ونسبة الاستفادة من الاجازات الإدارية وعدد أيام الاستراحات.

K	*K = 5*	*K = 4*	*K = 3*	*K = 4*	*K = 5*
	21	29	41	17	15
Gender	M	M	M	M	M
Certificate	High School	High School	Institute	Bachelor	Bachelor
Service Duration	10.6667	11.6207	10.7561	13.7647	13.3333
Absence Days Count	1.8095	1.7931	2.8049	4.2353	3.8667
Leave Utilization Rate	78.0952	79.6552	88.0488	92.9412	92.6667
Rest Days Count	8.3333	8.1034	12.0732	17.3529	15

الجدول 5.3 : توزع بيانات المجموعة (B) مع تغير قيمة المتحول (K).

- ظهرت مجموعة جديدة عند الانتقال إلى /k=5/ هي المجموعة /H/ والتي يغلب عليها العمال الذكور من الفئة المتعلمة – شهادة التعليم الأساسية أو الابتدائية– ممن بلغت سنوات الخدمة لديهم حوالي /17/ سنة بشكل متوسط. إن متوسط عدد أيام الغياب لهذه المجموعة هو يومان ونصف ويستهلكون حوالي 65% من استحقاق أيام الاجازات الخاصة بهم ومتوسط عدد أيام الاستراحات المرضية هو /10/ أيام.

من خلال إعادة التمعن بنتائج عملية العنقدة يمكن ملاحظة الانتقال من مرحلة الاحصاء الكلاسيكي القائم على الأرشفة كما يظهر في الملحق رقم (10) إلى مرحلة استكشاف المعرفة والمعلومات المفيدة, حيث يمكن لأصحاب القرار لدى التمعن في نتائج العنقدة – الشكل (4.3) – استخلاص ما يلي:

- مع تقدم عدد سنوات الخدمة يزداد التزام العاملين بأوقات الدوام الرسمي كما ينخفض مدى استفادتهم من أيام الغياب المشروع.
- العاملين الذكور أكثر التزاماً وتواجداً ضمن العمل.
- العاملين مما تتراوح سنوات عملهم ما بين 10 إلى 13 سنة هم أقل الفئات التزاماً بساعات العمل ويستهلكون أكثر من 90% من عدد أيام الغياب المشروع المجاز.

إن المعلومات التي تم استخلاصها محدودة بسبب محدودية البيانات المتواجدة ولكن في حال توافر بيانات لعدة سنوات وتضم عدة آلاف من العاملين لمختلف الفئات الوظيفية يمكن استنتاج معلومات أكثر دقة وأشمل من المعلومات الحالية.

الفصـــــــل الرابـــــــع

النتائـــج والتوصيـات

1.4. النتـــــــــــائج (Results)

شكلت هذه الدراسة رديفاً للعديد من للأبحاث الساعية للتأكيد على إمكانية الاستفادة من الأدوات المختلفة للتنقيب في المعطيات في الإدارة الذكية للموارد البشرية بالاستفادة من بيانات الفعاليات المتنوعة للكوادر البشرية.

تطابقت نتائج البحث مع الدراسات المرجعية التي أكدت على الدور الفعال للتنقيب في المعطيات على استكشاف المعرفة وخاصة مع وجود بيانات حقيقية, متوازنة وموثوقة. وفيما يلي نعرض أهم ما توصلت إليه الدراسة من نتائج:

1) تقدم أدوات التنقيب في المعطيات إمكانات واسعة لتطبيقات الذكاء الصنعي لشتى أنواع البيانات ومن بينها معطيات فعاليات الكوادر البشرية مما يسمح بتحليل متقدم لهذه البيانات يؤسس أطر وآليات الإدارة الذكية للموارد البشرية.
2) يشكل تطبيق Weka أداة فعالة للاستفادة من تقانات التنقيب في المعطيات من خلال ما يمنحه هذا التطبيق من خيارات وإمكانات متنوعة, حيث يسمح هذا التطبيق بتجهيز البيانات وفق صيغ متعددة قابلة للتعديل والحفظ والنقل بسهولة ومن ثم استخدامها وفق شتى تقانات التنقيب في المعطيات من التصنيف (Classification) مروراً بالتجزئة (Clustering) والربط (Association) وانتهاءً بانتقاء المعايير (Attribute Selection) بالإضافة لما يوفره هذا التطبيق من الإمكانات المرئية (Visualization) الأمر الذي يسمح بتسهيل العمل ومعاينة النتائج ومقارنتها بشكل أكثر فعالية. كما يمكن الاستفادة من هذا التطبيق ضمن البيئة البرمجية من خلال حزمة (Package) تمنح امكانية الاستفادة من جمع أدوات التنقيب في المعطيات بالتزامن مع الارتباط بكافة مكتبات لغة البرمجة Java بشكل فعال يفتح آفاقاً واسعة من الاستفادة والنتائج المثمرة الناتجة عن هذا الارتباط.
3) يشكل حسن اختيار بيانات التدريب أحد أهم عوامل نجاح عمليات التنقيب في المعطيات وخاصة بناء النماذج التنبؤية وتتجلى جودة بيانات التدريب في النقاط التالية:
    - تناسق البيانات وعدم وجود تعارض بين البيانات والفئات التي تنتمي إليها البيانات.

- تجنب القيم غير المنتظمة للبيانات بحيث تكون البيانات ضمن نطاق مناسب وبشكل خاص عندما يكون عدد بيانات التدريب محدوداً بالنسبة لعدد المعايير.
- التناسق بين عدد بيانات التدريب وعدد المعايير وقد ظهر هذا جلياً عند مقارنة نتائج دقة التصنيفات بين المنهجين /1/ و/2/, حيث كانت نتائج المنهج /2/ أكثر دقة وأكثر اعتدالاً وتقاربت دقة النتائج بين التصنيفات المختلفة.

4) باستخدام أداة التصنيف يمكن استخدام نتائج عملية تقييم أداء العاملين في بناء نماذج قادرة على التنبؤ بالوظيفة المناسبة للموظف بناء على تقييمه الحالي – وذلك مع الفرض بأن التقييم ذو مصداقية وشفافية وبعيد عن العلاقات والأهواء الشخصية.

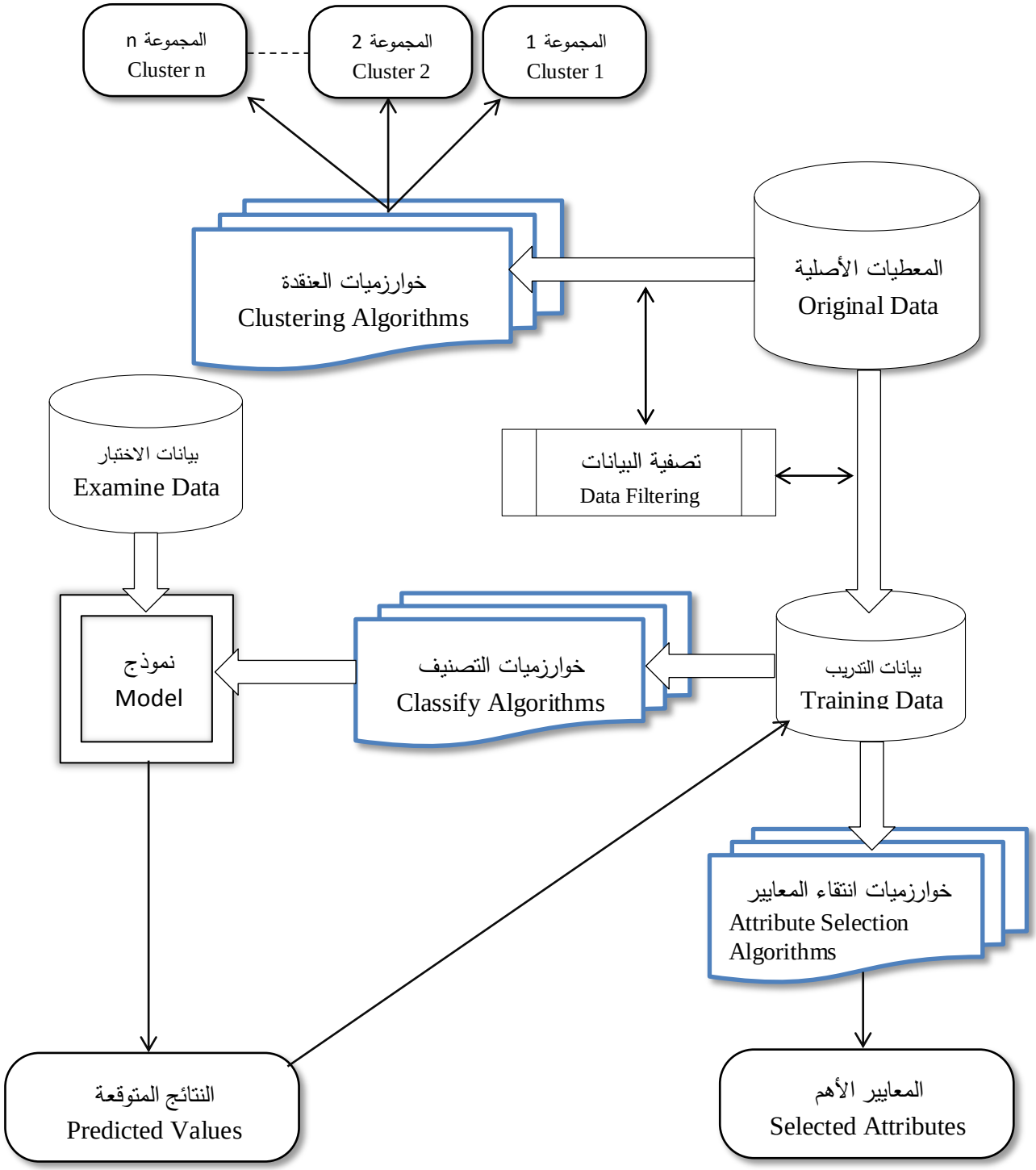
5) إن أداء كل من خوارزميات التصنيف (MLP, SVM, J48, Naïve Bayes) كان جيداً في التعامل مع هذا النوع من المعطيات بينما كان أداء خوارزمية (KNN) ضعيفاً مما يستدعي التوسع أكثر في دراسة مدى إمكانية الاستفادة من هذا المصنف للتعامل مع هذا النوع من البيانات.

6) تمنح خوارزميات انتقاء المعايير القدرة على تحديد أهم معايير التقييم التي يجب أخذها بعين الاعتبار ضمن فعاليات الكوادر البشرية المختلفة كالتوظيف والترشيح والترفيع وغيرها.

7) إن اختيار خوارزمية انتقاء المعايير مرتبط بالتطبيق العملي حيث تمنح خوارزمية (Ranker) انتقاءً سريعاً للمعايير بناء عملية التقييم (Evaluation) التي يتم اختيارها بينما تؤمن خوارزمية (Genetic) الحالة الأمثل العامة وخوارزمية (Greedy) الحالة الأمثل المحلية.

8) يتم تخزين بيانات التفقد اليومي للعاملين ضمن قواعد البيانات للاستفادة منها عبر طرق التحليل الكلاسيكية ولكن بالاستفادة من خوارزميات التجزئة تم تحليل بيانات عملية التفقد اليومي بشكل فعال لاستكشاف الروابط بين مجموعات العاملين بناء على هذه البيانات بعد إلحاقها بعدد من البيانات الأساسية المرتبطة بهم.

9) أظهرت النتائج الأداء الجيد لخوارزمية (Simple k-means) في عملية تجزئة هذا النوع من المعطيات مع التأكيد على اختيار عدد المجموعات المراد تجزئة البيانات إليها بشكل مناسب مع خصائصها كما يمكن إجراء ذلك بشكل تدريجي – بشكل مشابه لما تم القيام به ضمن هذا البحث.



الشكل 4.1 : النموذج المقترح للإدارة الذكية للموارد البشرية.

2.4. التوصيـــــــات (Recommendations)

توصي الدراسة بسبع توصيات رئيسية على ضوء النتائج التي تم عرضها سابقاً وذلك بهدف التطبيق العملي للنموذج المقترح بما يساهم في الإدارة الذكية للموارد البشرية بالاعتماد على نتائج عملية تقييم أداء العاملين وبيانات عملية التفقد اليومي:

1. إعطاء مفاهيم الذكاء الصنعي بشكل عام والتنقيب في المعطيات بشكل خاص أهمية خاصة ووضعها على رأس أولويات أعمال الكوادر المختصة بالأعمال الحاسوبية والمعلوماتية ضمن المؤسسات واجراء الدورات التدريبية لتأهيل تلك الكوادر للانتقال من المرحلة الكلاسيكية في جمع البيانات إلى مرحلة التحليل المتقدم للبيانات واستكشاف المعرفة.
2. منح عملية تقييم أداء العاملين أهمية أكبر من قبل المدراء والقائمين على عملية التقييم لما تحتويه نتائج هذا التقييم من معرفة مضمنة يمكن استخلاصها بالاستفادة من تقنيات التنقيب في المعطيات, هذا بالإضافة للتأكيد على إجراء عملية التقييم بشكل موضوعي وشفاف وتخطي العلاقات والروابط الشخصية أثناء عملية التقييم.
3. بناء نماذج تقييم موحدة على مستوى المؤسسات على الأقل مع مراجعة النماذج السابقة ومعالجة الإشكالات التي تعترض استثمارها ضمن تطبيقات الذكاء الصنعي لا سيما تداخل المعايير ووجود المعيار ذاته تحت عدة مسميات. كما يجب التنويه إلى إيلاء بعض المعايير أهمية خاصة – كأهداف الوظيفة – وذلك من خلال منحها أمثال وأوزان معينة تزيد من تأثيرها ضمن مجموعة المعايير.
4. توخي الدقة في جمع بيانات التفقد اليومي لما تحتويه هذه البيانات من معلومات مفيدة مرتبطة بالحالة المعنوية للعاملين.
5. اعتماد النموذج المقترح لمعالجة المعطيات الموضح في الشكل (4.1) والذي يتضمن الخطوات التالية:
   أ- جمع نتائج عمليات التقييم النصف سنوية بشكل تراكمي وجدولتها وتنظيمها.
   ب- جميع بيانات التفقد اليومي بشكل تراكمي وجدولتها وتنظيمها.

ت- معالجة البيانات وتصفيتها لتصبح جاهزة لعمليات التنقيب المختلفة (التصنيف, العنقدة وغيرها).

ث- اجراء عمليات التصنيف وفق الخوارزمية المتناسبة مع عدد بيانات التدريب مع الأخذ بعين الاعتبار أداء الخوارزميات التي تم اختبارها في هذا البحث.

ج- اعتماد مبدأ التصويت للأكثرية في اختيار الوظيفة المناسبة للموظف.

ح- في حال كان الإجماع بين جميع النماذج على الوظيفة التي تم اختيارها للموظف المطلوب يجب إضافة بيانات الاختبار إلى بيانات التدريب وبالتالي انجاز التعليم الذاتي (Self-Learning) للنموذج.

خ- تطبيق خوارزميات انتقاء المعايير بغية الوقوف على أهم معايير التقييم التي يجب مراعاتها عند انتقاء الموظفين الجدد أو اختيار موظفين لأماكن تتطلب مهارات خاصة.

د- اجراء تحليل متقدم لبيانات التفقد اليومي بالاستفادة من خوارزميات العنقدة مما يمنح نظرة أعمق لهذه البيانات وخاصة لدى دمجها مع بعض المعلومات الأساسية للموظف.

6. التوسع في هذه الدراسة من خلال الاستفادة من أدوات وخوارزميات أخرى للتنقيب في المعطيات لمعالجة هذا النوع من المعطيات وعلى وجه الخصوص استخدام خوارزميات هجينة ومقارنة نتائجها مع نتائج هذا البحث سعياً للوصول لأفضل الخوارزميات التي يمكنها التعامل مع هذا النوع من المعطيات.
7. التوسع في هذه الدراسة من خلال اعتماد معطيات ونتائج فعاليات أخرى للكوادر البشرية وإضافتها لمستودع البيانات لتنفيذ عمليات أوسع كالتنبؤ بالأداء المستقبلي أو انتقاء المواهب أو إدارة عمليات التدريب والتأهيل.

وأخيراً نذكر بأن الهدف الرئيسي لهذا العمل هو التقدم خطوة للأمام في تسلق هرم المعرفة الذي كان ولازال حلم البشرية الأول للوصول إلى الكمال.

المراجع (References)

1. المراجع العربية

1. أحسن, طيار, "**التنقيب في البيانات واتخاذ القرارات**", الملتقى الوطني السادس, جامعة سكيكدة, الجزائر, 27-28 كانون الثاني 2009.
2. أسماء عبد الغني قاسم الشربجي, 2006,"**اكتشاف أنماط هامة في بيانات الطلاب في جامعة العلوم والتكنولوجيا بتطبيق تقنيات التنقيب عن البيانات**", أطروحة ماجستير غير منشورة, كلية العلوم والهندسة, جامعة العلوم والتكنولوجيا, اليمن, 2006.
3. الأميري, "**تقييم وتطوير نظام تقويم أداء العاملين – دراسة حالة مصرف دجلة والفرات للتنمية والاستثمار**", رسالة ماجستير, الأكاديمية العربية المفتوحة, الدنمارك, 2009.
4. خلوف, رزوق, شميس, "**استخدام آلية التنقيب في المعطيات للمساعدة في اكتشاف عمليات الاحتيال في بيئة مصرفية**", مجلة جامعة دمشق للعلوم الهندسية المجلد الخامس والعشرون, العدد الثاني, 2009, جامعة دمشق.
5. خلوف, رزوق, شميس,"**تطوير آليات جديدة للتنقيب في المعطيات لإدارة علاقات الزبائن في بيئة مصرفية**", مجلة جامعة دمشق للعلوم الهندسية المجلد السادس والعشرون, العدد الاول, 2010, جامعة دمشق.
6. سلمان, "**التنقيب وإعادة هندسة قواعد البيانات**", المؤتمر السادس لجمعية المكتبات والمعلومات السعودية المنعقد في مدينة الرياض خلال الفترة 6-7 ابريل 2010م.
7. عمران, "**تقويم أداء العاملين في شركة الصناعات الإلكترونية سيرونيكس**", مجلة جامعة دمشق, المجلد 32, العدد الأول 2007.
8. المرسوم التشريعي /322/ لعام 2005, تاريخ 2005/7/27.
9. المنتدى العربي لإدارة الموارد البشرية, http://www.hrdiscussion.com/.
10. وزارة الشؤون الاجتماعية والعمل, لجنة تقييم أداء العاملين, نموذج تقييم الأداء الخاص بالوزارة, "**نموذج تقيم أداء الوظائف الإشرافية والقيادية**", "**نموذج تقيم أداء الوظائف الخدمية**", "**نموذج تقيم أداء الوظائف الإنتاجية**".
11. وزارة الشؤون الاجتماعية والعمل, لجنة تقييم أداء العاملين, "**نموذج التقييم 1/أ**", "**نموذج التقييم 1/ب**", "**نموذج التقييم 1/ج**".

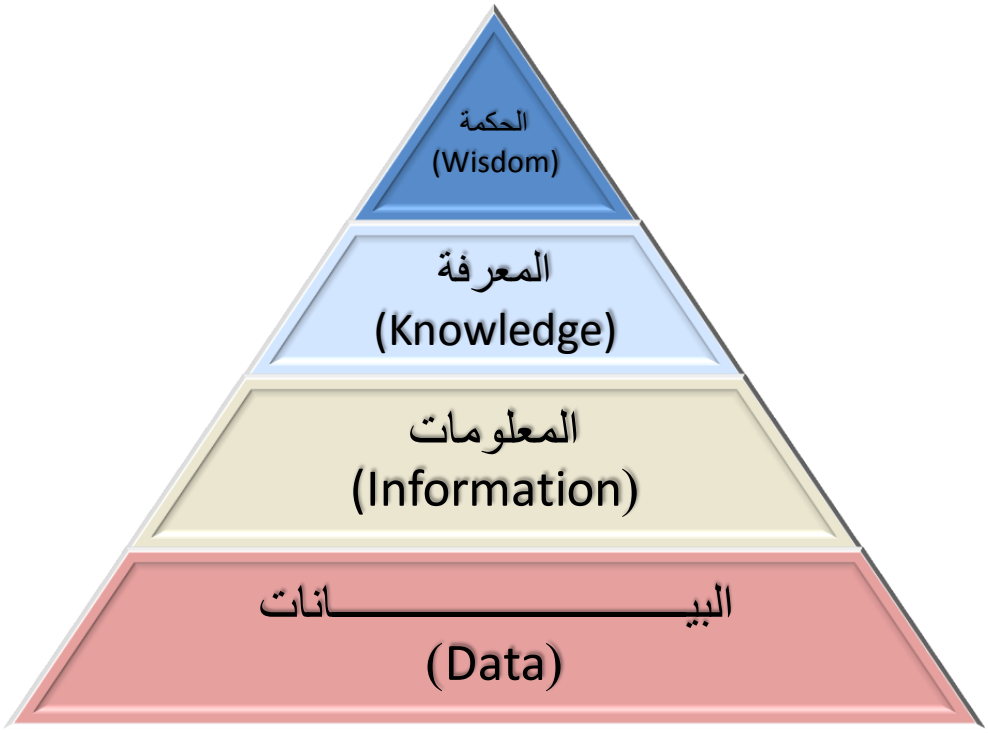
12.اليونس, 2007, "**تقييم أداء العاملين والترفيعات السنوية**", ورشة عمل أقامتها مديرية التدريب والتأهيل والمعلوماتية في المؤسسة العامة لسد الفرات, سورية.

2. المراجع الأجنبية

1. Jantan, Hamdan, Othman, "**Towards applying Data Mining Techniques for Talent Mangement**", 2009 International Conference on Computer Engineering and Applications IPCSIT vol.2 (2011) © (2011) IACSIT Press, Singapore.
2. Sadath, "**Data Mining: A Tool for Knowledge Management in Human Resource**", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-2, Issue-6, April 2013.
3. Edvardsson, "**HRM and knowledge management**", Literature review, Employee Relations Vol. 30 No. 5, 2008, pp. 553-561 ©Emerald Group Publishing Limited.
4. Goyal, Mehta, "**Performance Comparison of Naïve Bayes and J48 Classification Algorithms**", International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11 (2012), © Research India Publications; http://www.ripublication.com/ijaer.htm
5. Witten, Frank, "**Data Mining: Practical Machine Learning Tools and Techniques**", Second Ed., © 2005 by Elsevier Inc.
6. Mehta, Anjana, Iti, "**Reasoning with Missing Values in Multi Attribute Datasets**", Research Paper, *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 3, Issue 5, May 2013.
7. Bouckaert, Frank, Hall, Kirkby, Reutemann, Seewald, Scuse, "**WEKA Manual for Version 3-6-1**", June 4, 2009.
8. Naser, Masri, Abu Sultan, Zaqout, "**A PROTOTYPE DECISION SUPPORT SYSTEM FOR OPTIMIZING THE EFFECTIVENESS OF ELEARNINGIN EDUCATIONAL INSTITUTIONS**", *International Journal of Data Mining & Knowledge Management Process (IJDKP)* Vol.1, No.4, July 2011.
9. Kobbacy, Vadera, " **A survey of AI in operations management from 2005 to 2009**", Journal of Manufacturing Technology Management, Vol. 22 Iss: 6 pp. 706 – 733
10. Takeda, Helms, 2010, "**Globally sustainable management: a dynamic model of IHRM learning and control**", The Learning Organization, Vol. 17 Iss: 2 pp. 133 – 148.
11. Kuo, 2011, "**How to improve organizational performance through learning and knowledge?**", International Journal of Manpower, Vol. 32 Iss: 5 pp. 581 – 603.

12. Edvardsson, 2008, "**HRM and knowledge management**", Employee Relations, Vol. 30 Iss: 5 pp. 553 - 561.
13. Jantan, Hamdan, Othman, 2011, "**Data Mining Classification Techniques for Human Talent Forecasting"**, Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1, InTech, DOI: 10.5772/14007. Available from: http://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/data-mining-classification-techniques-for-human-talent-forecasting
14. Jantan, Hamdan, Othman, Puteh, 2010, "**Applying Data Mining Classification Techniques for Employee's Performance Prediction**", kmice.cms.net.my, http://www.kmice.cms.net.my/ProcKMICe/KMICe2010/Paper/PG601-607.pdf
15. Jantan, Hamdan, Othman, 2010, "**Intelligent Techniques for Decision Support System in Human Resource Management**" Decision Support Systems Advances in, Ger Devlin (Ed.), ISBN: 978-953-307-069-8, InTech, DOI: 10.5772/39401. Available from: http://www.intechopen.com/books/decision-support-systems-advances-in/intelligent-techniques-for-decision-support-system-in-human-resource-management
16. Kumar, Sahoo, 2012, "**Analysis of Bayes, Neural Network and Tree Classifier of Classification Technique in Data Mining using WEKA**", Academy & Industry Research Collaboration Center, http://airccj.org/CSCP/vol2/csit2236.pdf
17. Adel A., Mohammad V. S., Parviz A., P., Abdolreza R., (2013). **A model for personnel selection with a data mining approach: A case study in a commercial bank**, *SA Journal of Human Resource Management*, **11,** http://dx.doi.org/10.4102/sajhrm-.v11i1.449.
18.

الملحق رقم (1) : هرم المعرفة



تعاريف خاصة بهرم المعرفة[45]:

البيانات	المعلومات	المعرفة	الحكمة
رموز	بيانات تمت معالجتها لتصبح ذات فائدة	القدرة على الإجابة على التساؤلات حول "الكيف"	عملية تقييم لفهم الأمور
مجموعة غير مترابطة من الحقائق حول حدث معين	حقائق مبنية على معالجة البيانات ذات الفائدة والهدف المشترك للمساعدة في اتخاذ القرار.	مستوى عالي من الاستنتاجات والأفكار المتولدة في عقول الأفراد يضاف إليها المهارات والتدريب والفطرة السليمة والتجارب.	أعلى مستوى من الاستنتاجات مضاف إليها بصيرة الفرد وقدرته على رؤية ما وراء الأفق.

[45] بشرى الأفندي,2013, إدارة المعرفة التقانية في المؤسسات الإنتاجية (الواقع–العقبات) واقتراح نموذج لإدارة المعرفة التقانية في سورية بالتطبيق على شركة آسيا للصناعات الدوائية, أطروحة ماجستير, الجامعة الافتراضية السورية, ص 121.

الملحق رقم (2) : نماذج تقييم الأداء الخاصة بالشركة السورية للتبغ.

النموذج الخاص بالوظائف القيادية الإشرافية				
الكفاءات الأساسية	الفترة التي يغطيها التقرير			
	لغاية شباط	لغاية آب	لغاية شباط	لغاية آب
	التقدير	التقدير	التقدير	التقدير
يمتلك القدرة على التخطيط والتطوير				
يسعى الى تحسين أداء مرؤوسيه وخلق روح التعاون بينهم				
يمتلك القدرة على التعامل مع تقنيات الادارة الحديثة				
يحافظ على علاقات جيدة مع مرؤوسيه والآخرين				
يلتزم بالأنظمة والقوانين والتعليمات الناظمة لعمله وللدوام				
قدرته على تحمل المسؤوليات بشكل فعال				
ينمي قدراته الذاتية ويسعى لزيادة معرفته في مجال عمله				
يسعى لأداء عمله ضمن الوقت المحدد بدقة وبأقل أخطاء ممكنة				
يهتم بمتابعة قضايا الصحة والسلامة المهنية				
يمتلك القدرة على المبادرة والابتكار والابداع				
يتعرف على مشكلات العمل ويتخذ الاجراءات المناسبة لحلها				
يقدم لمرؤوسيه ما يحتاجونه لإنجاز عملهم بأفضل ما أمكن				
مهام الوظيفة	التقدير	التقدير	التقدير	التقدير
يشرف على تنفيذ التعليمات المتعلقة بسير عمله				
إجراء الدراسات والرد على الاستفسارات				
حسن توجيه مرؤوسيه وإشرافه على أعماهم				
يمارس صلاحياته بدقة ومسؤولية				
يقوم بالأعمال التي يكلف بها من قبل رؤسائه				
الأهداف	التقدير	التقدير	التقدير	التقدير
يطور ويحسن أداء العمل				
ينفذ برامج العمل الموضوعة لمكان عمله				
يرفع من كفاءات مرؤوسيه بالتدريب والتحفيز و..و.....				

الملحق رقم (2) : نماذج تقييم الأداء الخاصة بالشركة السورية للتبغ.

النموذج الخاص بالوظائف الانتاجية والخدمية				
الكفاءات الأساسية	الفترة التي يغطيها التقرير			
	لغاية شباط	لغاية آب	لغاية شباط	لغاية آب
	التقدير	التقدير	التقدير	التقدير
يحرص على الاستخدام الاقتصادي لمستلزمات العمل				
يمتلك روح الفريق بالعمل ويتعاون مع الأخرين لإنجاز المهام				
يتقيد بإجراءات الصحة والسلامة المهنية				
يلتزم بالدوام				
يهتم بتنمية قدراته الذاتية				
يتعامل باحترام مع رؤسائه وزملائه والمراجعين				
يؤدي عمله بتركيز وبأقل مستوى من الأخطاء				
يحدد مشكلات العمل ويقترح حلول مناسبة لها إن أمكنه ذلك				
يمتلك القدرة على التعامل مع التقنيات الحديثة				
ينجز مهامه بتكلفة منخفضة وجودة عالية وبالتوقيت المحدد				
يلتزم وينفذ القوانين والتعليمات وإجراءات العمل				
مهام الوظيفة	التقدير	التقدير	التقدير	التقدير
يسعى لتنفيذ برامج الإنتاج المحددة لعمله				
يحافظ على المنتج والمواد الداخلية فيه والتجهيزات من التلف				
يقوم بالأعمال التي يكلف بها من قبل رئيسه				
الأهداف	التقدير	التقدير	التقدير	التقدير
يطور ويحسن أداء العمل				
يسعى الى تطوير المنتج بأفضل ما يمكن				
يسعى الى تنفيذ سياسة الجودة				
يخفض نسب الهدر				
يرفع من وتيرة عمله ويحقق الغاية منه				

الملحق رقم (3) : نماذج تقييم الأداء الخاصة بالمؤسسة العامة لسد الفرات

النموذج الخاص بالوظائف الإشرافية والقيادية			
الكفاءات الأساسية	ملاحظات	التقدير	المعدل
القدرة على استشراف المستقبل والتنبؤ به وإيجاد الحلول للمشاكل الطارئة.			
القدرة على تحديد أولويات العمل وتحمل المسؤوليات واتخاذ القرارات المناسبة.			
المرونة في العمل ومنح السلطات والمسؤوليات للمرؤوسين.			
الالتزام بالقوانين والأنظمة والسعي لتطويرها والمحافظة على قواعد الصحة والسلامة المهنية.			
القدرة على المبادرة والابتكار واستثمار المعلوماتية في تطوير برامج العمل.			
الدوام والالتزام بمواعيده.			
تنمية قدراته ومبادراته الذاتية المتعلقة بعمله.			
علاقته بمرؤوسيه وخلق علاقات جيدة معهم.			
مهام الوظيفة	ملاحظات	التقدير	المعدل
الإشراف على حسن تطبيق القوانين والأنظمة المتعلقة بعمله.			
الإشراف على العمل والمرؤوسين للوصول إلى استثمار أمثل لإمكانياتهم.			
الإشراف على إعداد الدراسات والرد على الاستفسارات المتعلقة بالعمل وإعداد القرارات اللازمة.			
المشاركة بفعاليات التنظيمات النقابية لرفع سوية العمل والإنتاج.			
تنفيذ كافة المهام وفق متطلبات العمل.			
الأهداف	ملاحظات	التقدير	المعدل
المشاركة في تطوير القوانين والأنظمة والتعليمات المتعلقة بعمل المؤسسة.			
المشاركة في وضع التعليمات التنفيذية للقوانين المتعلقة بعمل المؤسسة (نظام داخلي, حوافز,الخ)			
إعداد الدراسات المتعلقة بتطوير أداء إدارات المؤسسة			
تقويم أداء مرؤوسيه تبعاً لعملهم وبحيادية مطلقة.			

الملحق رقم (3) : نماذج تقييم الأداء الخاصة بالمؤسسة العامة لسد الفرات

النموذج الخاص بالوظائف الفنية والخدمية			
الكفاءات الأساسية	ملاحظات	التقدير	المعدل
الإنتاجية واتقان العمل.			
الالتزام بقواعد الأمن الصناعي والصحة والسلامة المهنية.			
المحافظة على أدوات ومستلزمات الإنتاج.			
الالتزام بمواعيد الدوام والمواظبة على العمل.			
تحديد معوقات العمل والمساهمة في وضع الحلول لها.			
اللياقة وحسن المظهر والتعامل مع الرؤساء والزملاء والمراجعين.			
الالتزام بتطبيق القوانين والأنظمة المتعلقة بالعمل.			
تنمية القدرات الذاتية وتنفيذ المهام الموكلة له بالوقت المحدد.			
مهام الوظيفة	ملاحظات	التقدير	المعدل
تنفيذ برامج العمل في الوقت المحدد.			
تطبيق الشروط الفنية والقانونية للعمل وتأمين مستلزماته قبل الشروع به.			
إعداد المستندات ومسك السجلات اللازمة وإعداد الكتب والمراسلات المطلوبة.			
إعلام الرئيس المباشر عن أي طارئ عمل.			
تنفيذ كافة المهام التي يكلف بها من رئيسه المباشر وفق متطلبات العمل.			
الأهداف	ملاحظات	التقدير	المعدل
تنفيذ برامج الإنتاج والعمل المقررة وفق البرامج الموضوعة.			
تخفيض نسبة الهدر في المواد الاولية والعدد والآلات وقطع التبديل عن نسبة الهدر السابقة.			
زيادة وتيرة العمل عن الفترة السابقة.			
رفع مستوى الاداء واتباع الدورات التدريبية واستثمار الحاسب في العمل.			

الملحق رقم (4) : نماذج التقييم (أ, ب, ج) المصدقة من قبل وزارة الشؤون الاجتماعية والعمل.

النموذج الخاص بالوظائف الإشرافية والقيادية		
الكفاءات الأساسية	التقدير	ملاحظات
يمتلك القدرة على التنبؤ والتخطيط وجدولة العمل.		
يتعرف على مشكلات العمل ويتخذ القرارات المناسبة لحلها ويتحمل المسئولية بشكل فعال.		
يمنح المسؤولين السلطات والمسؤوليات الكافية لأداء أعمالهم.		
يمتلك القدرة على تحديد أولويات العمل بما يتلاءم مع أهميتها.		
يتابع باستمرار إنجاز المهام المطلوبة مع المرؤوسين.		
يتعامل مع الناس احترام ولطف ويقدم الإرشاد ويدعم الآخرين.		
يتواصل بشكل مفهوم وواضح عندما يصدر التوجيهات والتعليمات لمرؤوسيه.		
يشجع على التعاون وروح الفريق.		
يحافظ على علاقات جيدة مع الرؤساء والمرؤوسين.		
يلتزم بقوانين وأنظمة وقواعد وإجراءات العمل وسياسة المنظمة وينفذها ويسعى لتطويرها.		
يمتلك القدرة على التركيز بالعمل والمحافظة على علاقات بناءة مع الأفراد الآخرين في ظروف ضغوط العمل.		
يسعى لأداء العمل ضمن حدود الوقت المحدد بدقة وبأقل مستوى أخطاء ممكن.		
يمتلك القدرة والمهارات لتطوير أداء المرؤوسين وتحفيز العاملين باستخدام جميع أنواع الحوافز.		
يمتلك القدرة اللازمة للتعامل مع تقنيات الإدارة الحديثة والقدرة على المبادرة والابتكار والإبداع.		
يلتزم بالمواظبة على الدوام وأوقات الحضور والانصراف.		
مهام الوظيفة	التقدير	ملاحظات
تنفيذ الأهداف المحددة والمخططة من قبل الإدارة العليا بدقة في مواعيدها وإيضاحها غلى المستوى الأدنى وإعطاء التوجيهات.		
المحافظة على النظام والانضباط ورفع التقارير والبيانات عن الأعمال المنفذة إلى الجهات الأعلى.		
إبداء الرأي وتقديم المقترحات حول تحسين العمل ومكافحة الهدر والإسراف.		
توزيع المهام والأعباء على العاملين والاهتمام بالسلامة العامة.		
الأهداف	التقدير	ملاحظات
تنفيذ القوانين والمراسيم والقرارات والأنظمة الخاصة بالوزارة.		
العمل على تطوير كل ما من شأنه تحسين العمل ومكافحة الهدر والإسراف.		

الملحق رقم (4) : نماذج التقييم (أ, ب, ج) المصدقة من قبل وزارة الشؤون الاجتماعية والعمل.

النموذج الخاص بالوظائف الخدمية والفنية		
الكفاءات الأساسية	التقدير	ملاحظات
يمتلك القدرة على التعامل مع المراجعين بلطف واحترام وبشكل واضح ومفهوم.		
يحافظ على حسن المظهر ويلتزم باللباس الرسمي في حال وجوده.		
يحافظ على نظافة مكان العمل وترتيبه ليظهر بشكل لائق أمام الآخرين.		
يؤمن بأن العميل على حق دائماً.		
يتقيد بإجراءات وبرامج الصحة والسلامة المهنية من خلال التأكد من سلامة التجهيزات والمحافظة عليها.		
يحدد مشكلات العمل ويقترح حلول مناسبة لها كل ما أمكنه ذلك.		
يحرص على الاستخدام الاقتصادي لمستلزمات العمل المادية في مجال وظيفته وبالوقت المحدد.		
يمتلك القدرة على تحديد أولويات مهام العمل حسب أهميتها.		
يحافظ على علاقات جيدة مع رؤسائه ويتعامل بلطف واحترام مع زملائه بالعمل.		
يتعاون مع الآخرين لإنجاز المهام المشتركة ويمتلك روح الفريق في العمل.		
يلتزم وينفذ القوانين وأنظمة وقواعد وسياسات وإجراءات العمل.		
يحافظ على التركيز بالعمل وعلى علاقات بناة مع الآخرين حتى في حال تعرضه لضغوط العمل.		
يسعى لأداء المهام ضمن الوقت المحدد بدقة وبأقل مستوى من الأخطاء		
يهتم بتنمية قدراته الذاتية		
مهام الوظيفة	التقدير	ملاحظات
تسيير أعمال الإدارة وخدمة أهدافها.		
مساعدة الإدارة في أداء واجباتها.		
الأهداف	التقدير	ملاحظات
ضمان تنفيذ الأعمال الخدمية والفنية بدقة ووضوح.		

الملحق رقم (5) : مثال تطبيقي للاستفادة من الحزمة البرمجية الخاصة بتطبيق WEKA

```
import weka.classifiers.*;
import weka.classifiers.trees.*;
import weka.experiment.*;
import weka.core.*;
import weka.classifiers.Classifier;
import weka.classifiers.bayes.NaiveBayes;
import weka.classifiers.trees.J48;
import weka.attributeSelection.*;
import weka.gui.*;
import weka.associations.*;
import weka.clusterers.*;
import weka.classifiers.Evaluation;
import weka.core.Attribute;
import weka.core.Instance;
import weka.core.Instance;
import weka.core.Instances;

public static void main(String[] args) throws Exception{

      Attribute Attribute1 = new Attribute("Knowledge");
      Attribute Attribute2 = new Attribute("Execution");
      Attribute Attribute3 = new Attribute("Problem-Solving-and-Resolution");
      Attribute Attribute4 = new Attribute("Process-Improvement");
      Attribute Attribute5 = new Attribute("Safety");
      Attribute Attribute6 = new Attribute("Productivity");
      Attribute Attribute7 = new Attribute("Quality");
      Attribute Attribute8 = new Attribute("Initiative");
      Attribute Attribute9 = new Attribute("Attendance-and-Punctuality");
      Attribute Attribute10 = new Attribute("Organization");
      Attribute Attribute11 = new Attribute("Adaptability");
      Attribute Attribute12 = new Attribute("Politeness");
      Attribute Attribute13 = new Attribute("Personal-Appearance");
      Attribute Attribute14 = new Attribute("Technical-Writing");
      Attribute Attribute15 = new Attribute("Presentations");
      Attribute Attribute16 = new Attribute("Listening");
      Attribute Attribute17 = new Attribute("Negotiation_and_Influence");
      Attribute Attribute18 = new Attribute("Facilitation");
      Attribute Attribute19 = new Attribute("Responding-to-Conflict");
      Attribute Attribute20 = new Attribute("Interaction-with-Coworkers");
      Attribute Attribute21 = new Attribute("Interaction-with-Supervisors");
      Attribute Attribute22 = new Attribute("Interaction-with-E/C");
      Attribute Attribute23 = new Attribute("Motivational-Skills");
      FastVector fvClassVal = new FastVector(4);
        fvClassVal.addElement("A");
        fvClassVal.addElement("B");
        fvClassVal.addElement("C");
        fvClassVal.addElement("D");
      Attribute ClassAttribute = new Attribute("Result", fvClassVal);
      FastVector fvWekaAttributes = new FastVector(24);
        fvWekaAttributes.addElement(Attribute1);
        fvWekaAttributes.addElement(Attribute2);
        fvWekaAttributes.addElement(Attribute3);
        fvWekaAttributes.addElement(Attribute4);
        fvWekaAttributes.addElement(Attribute5);
        fvWekaAttributes.addElement(Attribute6);
        fvWekaAttributes.addElement(Attribute7);
        fvWekaAttributes.addElement(Attribute8);
        fvWekaAttributes.addElement(Attribute9);
        fvWekaAttributes.addElement(Attribute10);
```

```
        fvWekaAttributes.addElement(Attribute11);
        fvWekaAttributes.addElement(Attribute12);
        fvWekaAttributes.addElement(Attribute13);
        fvWekaAttributes.addElement(Attribute14);
        fvWekaAttributes.addElement(Attribute15);
        fvWekaAttributes.addElement(Attribute16);
        fvWekaAttributes.addElement(Attribute17);
        fvWekaAttributes.addElement(Attribute18);
        fvWekaAttributes.addElement(Attribute19);
        fvWekaAttributes.addElement(Attribute20);
        fvWekaAttributes.addElement(Attribute21);
        fvWekaAttributes.addElement(Attribute22);
        fvWekaAttributes.addElement(Attribute23);
        fvWekaAttributes.addElement(ClassAttribute);

        // Create an empty training set
        Instances isTrainingSet = new Instances("Rel", fvWekaAttributes, 1000);

        // Set class index
        isTrainingSet.setClassIndex(23);
        Instance iExample = new Instance(24);


        while ( rs.next() ){
             int k = 0;
              int j = 2;
              for ( ; j < numberOfColumns; j++, k++ ){
                   iExample.setValue((Attribute)fvWekaAttributes.elementAt(k),
                   (float) (rs.getFloat( j )));
                   }

               iExample.setValue((Attribute)fvWekaAttributes.elementAt(k),(String)
               (rs.getString( j )));
               isTrainingSet.add(iExample);
              }
              // NaiveBayes
         Classifier cModel = (Classifier) new NaiveBayes();
         cModel.buildClassifier(isTrainingSet);
         Evaluation eTest = new Evaluation(isTrainingSet);
         eTest.evaluateModel(cModel, isTrainingSet);
         String strSummary = eTest.toSummaryString();
          System.out.println(strSummary);
         double[][] cmMatrix = eTest.confusionMatrix();
         for(int row_i=0; row_i<cmMatrix.length; row_i++){
            for(int col_i=0; col_i<cmMatrix.length; col_i++){
                  System.out.print(cmMatrix[row_i][col_i]);
                  System.out.print("|");
                   }
                  System.out.println();
             }
           // J48
         cModel = (Classifier) new J48();
         cModel.buildClassifier(isTrainingSet);
         eTest = new Evaluation(isTrainingSet);
         eTest.evaluateModel(cModel, isTrainingSet);
         strSummary = eTest.toSummaryString();
        System.out.println(strSummary);
        cmMatrix = eTest.confusionMatrix();
         for(int row_i=0; row_i<cmMatrix.length; row_i++){
             for(int col_i=0; col_i<cmMatrix.length; col_i++){
                   System.out.print(cmMatrix[row_i][col_i]);
                   System.out.print("|");
```

```
                    }
                    System.out.println();
                }

                Instances unlabeled = new Instances(
                new BufferedReader(
                new FileReader("//SQL_Predict_1.arff")));
                unlabeled.setClassIndex(unlabeled.numAttributes() - 1);
                Instances labeled = new Instances(unlabeled);
                for (int i = 0; i < unlabeled.numInstances(); i++) {
                    double clsLabel =
                    cModel.classifyInstance(unlabeled.instance(i));
                        labeled.instance(i).setClassValue(clsLabel);
                    }
                BufferedWriter writer = new BufferedWriter(
                new FileWriter("//labeled.arff"));
               writer.write(labeled.toString());
               writer.newLine();
               writer.flush();
               writer.close();
              System.out.println("Training finished!");

               }
               catch(Exception e) {
                        e.printStackTrace();
                    }
           finally {
              if (rs != null) try { rs.close(); } catch(Exception e) {}
                  if (cstmt != null) try { cstmt.close(); } catch(Exception e) {}
                  if (con != null) try { con.close(); } catch(Exception e) {}
                  System.exit(1);
                  }
}
```

الملحق رقم (6) : بيانات التدريب بصيغة ملف ARFF حسب المنهج /1/

يتضمن بيانات التدريب التي تم تنفيذها بشكل عملي في هذا البحث حسب المنهج /1/ المتضمن /18/ معياراً مختلفاً. تم ترتيب البيانات حسب الفئات الوظيفية المعتمدة في هذا البحث (A, B, F).

```
@relation Evaluation
@attribute Knowledge numeric
@attribute Execution numeric
@attribute Problem-Solving-and-Resolution numeric
@attribute Self-Development numeric
@attribute Safety numeric
@attribute Productivity numeric
@attribute Quality numeric
@attribute Initiative-and-Inovation numeric
@attribute Attendance-and-Punctuality numeric
@attribute Organization numeric
@attribute Adaptability numeric
@attribute Politeness numeric
@attribute Personal-Appearance numeric
@attribute Economic numeric
@attribute Strategic-Thinking numeric
@attribute Interaction-with-Coworkers numeric
@attribute Interaction-with-Supervisors numeric
@attribute Interaction-with-E/C numeric
@attribute ServiceFunctions numeric
@attribute Org.Objectives numeric
@attribute Result {A,B,F}
@data
4.9,4.9,4.9,5,4.9,5,4.9,4.8,5,5,5,5,5,4.5,5,4.8,5,4.8,5,5,A
4.8,4.8,4.6,5,5,5,4.8,4.8,5,4.8,4.8,5,5,4.8,5,5,5,4.8,5,5,A
4.8,4.6,4.6,4.5,5,5,4.8,4.8,5,4.8,4.8,5,4.8,4.8,5,4.5,4.8,4.8,4.8,5,A
4.7,4.5,4.3,4.4,4.3,4,4.8,4.1,4.8,4.3,4.5,4.6,4.8,4.5,4.5,4.3,4.6,4.3,4.8,4.8,A
4.8,4.9,4.8,5,5,5,4.8,4.8,5,5,5,5,5,4.8,5,4.5,4.6,4.8,5,5,A
4.8,4.7,4.6,4.8,4.8,4.6,4.7,4.6,5,5,5,5,5,4.8,5,4.8,5,4.7,4.5,4.5,A
4,4,4,3.9,4.5,3.9,3.9,3.9,4,4.4,4.4,5,4.8,4.6,4.2,4.7,4.7,4.8,4.8,4.8,A
4.5,4.6,4.2,3.9,4.5,4.3,4.3,4,4.6,4.7,4.7,4.8,4.8,4.8,4,4.3,4.5,4.3,4.8,4.8,A
4.6,4.4,4.2,4.2,4.4,4.2,4.2,4,4.5,4.3,4.5,4.3,4.5,4.2,4,4.5,4.5,4.2,4.8,4.8,A
4.6,4.6,4.3,4.4,4.4,4.4,4.4,4.4,4.2,4.5,4.5,4.5,4.9,4.6,4.2,4,4.5,4.6,4.7,4.9,4.8,A
4.7,4.6,4.7,4.6,4.3,4.3,4.3,4.3,4.5,4.5,4.6,4.8,4.7,4.4,4.4,4.4,4.3,4.3,4.3,4.8,4.8,A
4.7,4.4,4.4,4.1,4.2,4.2,4.2,4.1,4.2,4.1,4.4,4.8,4.6,4.2,4,4.6,4.7,4.8,4.5,4.5,A
4.2,4.2,4.4,4,4.4,4.4,4,4.3,8,4,4,4,3.8,4,3.8,4.4,3.8,4.5,4.5,4,4,A
4.8,4.6,4.4,4.4,4.4,4.5,4.2,4.2,4.2,4.8,4.8,4.6,5,4.9,4.6,4,4.7,4.8,4.9,4.5,4.8,A
4.7,4.7,4.5,4.6,4.6,4.7,4.7,4.6,4.2,4.3,4.4,4.6,4.7,4.2,4.4,4.6,4.6,4.6,4.4,4.5,A
4.2,4.3,4.3,4.3,4.3,4.3,4.3,4,4,4,4,4.8,4.4,4.3,4.4,4.4,4.6,4.6,4.8,4.7,A
4.3,4.3,4.3,4,4.2,4.2,4.3,3.8,4,3.8,4,4.4,4.1,4,3.8,4,4,4,4.2,4.3,A
4.2,4.1,3.9,3.9,4,4.1,4.1,3.8,4.2,4,3.8,4.4,4,4,3.7,4,3.8,4,4.5,4.5,A
4.2,4.5,4.5,4.4,4.4,4.4,4.8,4.3,4.1,4.4,4.3,3.8,4.9,4.4,4,4,4.2,4.3,4.2,4.5,4.5,A
4.1,4.1,4.1,4.2,4.3,4.4,4.5,4.5,4.6,4.4,4.2,4.8,4.7,4,4.5,4.7,4.7,4.8,4.5,4.5,A
4.3,4.4,4.5,4.8,4.6,4.7,4.7,4.8,4.8,4.7,4.6,5,4.8,4.7,4.4,4.4,4.4,4.4,4.3,4.3,A
4.1,4.4,4.5,4.4,4.6,4.7,4.7,4.8,4.7,4.7,4.6,5,4.8,4.7,4.2,4.6,4.6,4.6,4.3,4.3,A
4.4,4.2,4,4.4,4.3,4.3,4.3,4.2,4.8,4.7,4.5,4.8,4.6,4.6,4.2,4.2,4.2,4.2,4.3,4.4,A
4.5,4.5,4.4,4.2,4.5,4.5,4.4,4.4,4.7,4.3,4.5,4.6,4.5,4.5,4.5,4.3,4.4,4.5,4.4,4.5,A
```

4.5,4.5,4.5,4.5,4.2,4.3,4.3,4.4,4.7,4.3,4.4,4.5,4.5,4.5,4.5,4.3,4.4,4.5,4.4,4.5,A
4.5,4.3,4.2,4,4.3,4.2,4.2,4,4,4,4.2,4.3,4.3,4.5,4.4,4.5,4.5,4.5,4.4,4.4,A
4.5,4.5,4.4,4.4,4.4,4.4,4,4,4,4.4,4,4,4.7,4.7,4.4,4,4.4,4.6,4.6,4.4,4.4,A
4.7,4.4,4.3,4.8,4.7,4.3,4.3,4.3,4.8,4.6,4.5,4.8,4.8,4.8,4.3,4.4,4.5,4.5,4.5,4.5,A
4.2,4.2,4,4.3,4.2,4.1,4.1,4,4.5,4.4,4.4,4.4,4.4,4.4,4.3,4,4.2,4.4,4.3,4.2,4.3,A
4.2,4.3,4.1,4.5,4.3,4.2,4.2,4,4.4,4.3,4.2,4.2,4.1,4.2,4,4,4.3,4.3,4.3,4.3,A
4.3,4.3,4.4,4.5,4.5,4,4.2,4.2,4.2,4.3,4.5,4.8,4.8,4.2,4.2,4.2,4.6,4.6,4.4,4.6,A
4.4,4.4,4.5,4.2,4.5,4.5,4.6,4,4.8,4.5,4.8,4.8,4.7,4.2,4.1,4.6,4.7,4.8,4.8,4.8,A
4.5,4,3.7,4.2,4.5,4,4,4,4.2,4,4.2,4.7,4.1,4.4,4.3,4.2,4.4,4.4,4.2,4.2,A
4.4,4.4,4.5,4,4.5,4.3,4.2,4,4,4,4.2,4,4.4,4.1,4.2,4.1,4.2,4.4,4.4,4.2,4.3,A
4.3,4,3.8,4.4,4.4,4.3,4.2,4,4.6,4.5,4.5,4.8,4.6,4.3,4,4.7,4.8,4.8,4.4,4.6,A
4.5,4.2,4.5,4.3,4.4,4.5,4,4.2,4,4.5,4.3,4,4,4.2,4.2,4.2,4.3,4,4.3,4.2,A
4,4.3,4.3,4,4.2,4.3,4,3.8,4.3,4.2,4.4,4.8,4,4.2,4.2,4.4,4.4,4.6,4.3,4.3,A
4.2,4.3,4.4,4,4,4,4.4,4,4,4,4.4,4,4,4.4,4.2,4,4.1,4,4.1,4.4,4.3,4.4,A
4.4,4.4,4.2,4.1,4.2,4.3,4.1,4.1,4.4,4,4.1,4.5,4.1,4.2,4.2,4.1,4.1,4.3,4.3,4.3,A
4.5,4,4,4,3.8,3.9,3.8,4,3.8,3.8,4.2,4.8,4.8,4,4,4.1,4,4.4,3.8,3.9,B
3.5,3.8,3.6,3.8,4,4.2,3.8,3.6,3.7,3.8,4.2,4.4,4.4,3.6,3.6,3.8,4,4,3.8,3.8,B
3.7,4,4,3.8,4.2,4,4,3.9,4,4.2,4.2,4.6,4.6,4.4,4.2,4.2,4.2,4.2,3.8,3.9,B
3.8,3.8,3.5,3.7,3,3.1,3.1,2.8,3,3,3.5,4.5,4.3,4,3.5,4,4.2,4,3.8,3.7,B
4.2,4.1,4,4,4.2,4.1,4,3.8,4.3,4.2,4.3,4.3,4.1,4,4,4,4,4,4,4,B
3.5,5,5,3.5,5,5,5,5,5,3.5,5,5,5,5,5,4,4,4,4,4,B
3.8,4,4.2,4.2,4.2,4.3,4,3.6,3,3.5,4,4.3,3.6,3,3,4,4.5,4,3.5,3.5,B
3.5,3.8,4,3.8,4,4,4,3,3.5,3,4,4.2,3.5,2.5,3,4,4,4,3.7,3.8,B
4,4,4.5,3.5,4,4.2,4,3.8,3.8,3.2,3,4.5,4,3,3,4.1,4.5,3.8,4,4,B
4,4,4,4,4,3.8,3.5,4.2,4.5,4.5,4.5,4.5,4.5,4,3.8,4,4.2,4.5,4.1,4.2,B
4,3,3,3,4,4,3.5,4,4,4,4,4,4,3,3.5,3,4,4,3.7,3.5,B
4.5,4.5,3.5,3,4.5,4,4,4,4.5,4,3,4.5,4.5,4,4,4,4,4,3.2,3,B
4,3.5,3,3,4,4,4,3.5,4,4,4.5,4.8,4.5,4.5,4,4.5,4.5,4.5,3,3,B
4,3.5,4,3.5,4,4,3,3,4,4,3,4,4,4,4,3.5,4,3.5,4,4,B
4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4.1,4.1,B
4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,3.4,3.2,B
4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,3.5,3.5,B
4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,B
4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,3.5,3.5,B
3.5,3.5,3,3,4,3.5,3,2,4,3.5,3,4,4,3,3,4,4,4,3.6,3.5,B
3.5,3.5,3,3.3,4,4,4,3.5,4,3,3.5,4,4.3,4,4,3,4,4,4,3.9,4,B
3.8,3.9,3.5,3.7,4,4,4,4,4.3,4,4,4.8,4.4,4,3.5,4,4,4,4,4,B
4,4,4,4.3,4,4,4.5,4.5,4,4.3,4,3.4,3,4.3,4.7,4.6,4,4,4.2,4.2,4.2,4,4,B
4,4,3.5,3.5,4.3,3.5,4.5,3.5,4.4,4.2,3.5,4.5,4.5,4,3.5,4,4,4,4,4,B
4,4,4.3,4.5,4,4,4,4,4.2,4,3.5,4.3,4.2,4,4,4,4,4,3.6,3.8,B
3.7,3.8,3.6,4,4,4,4,3.5,4.3,4,4.2,4.4,4.2,4,4,4,4.2,4.2,3.8,3.8,B
3.5,3.5,3.7,3.7,4,3.8,3.8,3.7,3.8,3.8,4,4.3,4.3,4,3.5,3.9,4.3,3.8,4,4,B
4,4,3.8,4,4,4,4,3.8,3.8,3.9,4,4.3,4.2,4,4,4,4.2,4.2,3.9,3.9,B
4.2,3.8,4,3.8,4,3.8,3.8,3.5,4.3,4,4.2,4.2,4.3,4,3.8,3.8,4,4.2,3.9,3.9,B
4,4,3.7,3.8,4,3.8,3.4,3,4,3.5,3.7,4.2,4.3,4,3.5,4,4,4,3.2,3.4,B
4,4,4,4,4,4.2,4.2,4,4.2,4,3.8,4,4,4,4,4.2,4,4,3.8,3.9,B
4.2,4,3.7,4,4,3.8,3.8,3.5,4,4.2,4,4.9,4.8,4.4,3.5,4,4.1,4,4,4,B
4,3.7,3.5,3.5,4,4,3.7,3,3,3,3.9,4.4,4,4,4,3,4.2,4.3,4.3,3.8,3.9,B
4,3.8,3.5,3,4,4,4,3,4,4,4,4.2,4.2,4,3,3.5,4,3.5,3.5,3.5,3.7,B
4,4,3.5,3,3.5,3.5,3.5,3.5,3.5,3.5,3.5,4,4,4,4,4,4,4.5,4,3.6,3.8,B
4,4,4.3,4.4,4,4.2,4.2,3.5,4,4,4.2,4.3,4,4,4.2,3.9,4.2,4,4.2,4,B
4,4,3.5,4,4.2,4.2,3.5,3.7,3.8,3.6,4,4.2,4.5,3.6,3.6,4.2,4.4,4.4,3.8,3.7,B
4,4,4.2,4.2,4.2,4.4,4,4,4,3.8,4.3,4.7,4.7,4,4,4.7,4.6,4.8,4,4,B

3.8,3.5,3.5,4,3.3,3.8,3.8,3.8,3.5,3.5,4,4.4,4.9,4,3.5,4.5,4.7,4.5,4,4,B
4.2,4.2,4.2,4,4,4,4.2,3.7,4.3,4.5,4.2,4.8,4.7,4.4,4,4.2,4.4,4.3,4.3,4.3,B
4,4,3.8,4,4,4.1,3.9,3.5,3.6,3.8,3.9,4.4,4.4,4,4,4,4,4,4,4,B
4.2,3.8,4.2,4,4.2,4,3.8,4.1,4.2,4.2,4,4.3,4.4,4.2,3.8,4.1,4.3,4.2,4,4,B
4,3.8,3.8,3.9,4,3.9,3.6,3.5,4,3.6,4,4.3,4.1,4,3.7,4.2,4.2,4,3.9,3.8,B
3.6,4,3.8,3.6,4,4,4,3.8,4.2,4,3.8,4,4,3.8,3.7,4,4.2,4.2,4,4,B
3.5,2,2,2,3,2,2,2,3.5,2,3,3,2.5,3,2,3,3,3,2,2,F
3,2,3,2.5,3,3,2,2,3,2,3,4,4,4,2,2,3,2,2.5,2.5,F
4,4,3.5,4,4,3.5,3.7,3,4,4,3,2,4,4,3.8,2,3,2,2.2,2.2,F
3.5,3,3,2.5,4,2,2,2,2,3,2.5,4,4,4,3,3,4,3,2,2.5,F
3,2,2,2,2,2,2,2,3,2,2,3.5,3,2,2,3,4,2,2,2,F
4,2,3,3,3,3,3,2,4,4,4,3.5,4,4,3,3,3,3,2.5,3,F
3.5,3,2,2,3,2,2,2,3,4,3,3,3.5,3,3,2.5,3.5,2,2,2.5,F
3.7,4,3,2,4,2,2,2,3.5,4,3,3,4,3,3,3,3,2,2,2,F
3,2,3,2,3,2,2,2,2,2,3,3,3,3,4,3,4,2,2,2,F
3,3,2,3,4,3,2,2,2,2,2,2,4,3,3,3,4,3,2.2,2.4,F
4,4,4,4,4,3,3,2,3.5,4,3.5,2,4,3,4,3,4,2,2,2,F
4.8,4.5,4.3,4,4,4,4,4,4,4.5,4.5,4.5,4.5,4.8,4,4,4.5,4.5,4.5,3,3,F
4,2,2,2.5,3,2,2,2,3.5,3.5,3,2.5,3,3,4,2.5,3.5,3,2,2,F
2,2,2.5,2,3,2,2,2,2.5,2,3.5,2,2,3,4,2,3.5,2,2,2.5,F
3,3,2,2.5,3,3,3,2,3,2,3,3.5,3,3,3,3,3.5,3,2.5,2,F
4,2.5,3.5,4,3,2.5,2.5,3,3.5,2,4,3,3,3,3.5,3,3,3,2.5,2.5,F
3.5,3,3,2,3,2,2,2,3,2.5,3.5,4,4,3,3,4,4,4,2.5,2.5,F

الملحق (7) – بيانات الاختبار بصيغة ملف ARFF حسب المنهج /1/

يتضمن بيانات الاختبار التي تم تنفيذها بشكل عملي في هذا البحث حسب المنهج /1/ المتضمن /18/ معياراً مختلفاً. تم ترتيب البيانات حسب انتظامها حيث ضمت بيانات الاختبار الست الأولى بيانات منتظمة أما البيانات المتبقية فهي بيانات اختبار غير منظمة.

```
@relation Evaluation
@attribute Knowledge numeric
@attribute Execution numeric
@attribute Problem-Solving-and-Resolution numeric
@attribute Self-Development numeric
@attribute Safety numeric
@attribute Productivity numeric
@attribute Quality numeric
@attribute Initiative-and-Inovation numeric
@attribute Attendance-and-Punctuality numeric
@attribute Organization numeric
@attribute Adaptability numeric
@attribute Politeness numeric
@attribute Personal-Appearance numeric
@attribute Economic numeric
@attribute Strategic-Thinking numeric
@attribute Interaction-with-Coworkers numeric
@attribute Interaction-with-Supervisors numeric
@attribute Interaction-with-E/C numeric
@attribute ServiceFunctions numeric
@attribute Org.Objectives numeric
@attribute Result {A,B,F}
@data
5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,?
4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,?
3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,?
2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,?
4.2,4.4,4.1,4,4,4,4.1,4.2,4.2,4.1,4,4.1,4,4,4,4,4.1,4,4,4,?
3.5,3.2,3.8,3.9,3.1,3.9,3.7,4,3.8,4,4,4,4,3.2,5,4.2,5,4,3.8,3.2,?
3,3.8,4,3.3,3.5,3.9,3.9,3.9,4,4.4,4.4,3,3.8,3.6,3.2,3.7,3.7,3.8,3.8,3.8,?
3.8,4.6,4.2,3.9,3.5,3.3,4.1,4,4.2,4.1,4.2,4.2,4.1,3.8,4,3.3,4,4.1,3.8,4.2,?
3,4,5,2,3,4,5,2,3,4,5,3,2,3,4,5,4,4,4,4,?
4,4,5,4,4,4,5,2,4,4,5,4,2,4,4,5,4.3,4.3,3.5,3.5,?
3,4,5,2,3,4,3.5,2,3,4,3.5,3,2,3,3.5,5,3,3,4.5,4.5,?
2,5,5,2,5,2,5,2,5,2,5,3,4,3,3.5,5,3,3,2,5,?
```

الملحق (8) – بيانات التدريب بصيغة ملف ARFF حسب المنهج /2/

يتضمن التالي بيانات التدريب التي تم تنفيذها بشكل عملي في هذا البحث حسب المنهج /2/ المتضمن /4/ معايير مختلفة. تم ترتيب البيانات حسب الفئات الوظيفية المعتمدة في هذا البحث (A, B, F).

```
@relation Evaluation
@attribute Basic-Skills numeric
@attribute Service.Functions numeric
@attribute Org.Objectives numeric
@attribute Result {A,B,F}
@data
4.9,5,5,A 4.9,5,5,A  4.8,4.8,5,A 4.5,4.8,4.8,A 4.9,5,5,A 4.8,4.5,4.5,A 4.3,4.8,4.8,A 4.4,4.8,4.8,A 4.3,4.8,4.8,A
4.5,4.9,4.8,A 4.5,4.8,4.8,A 4.4,4.5,4.5,A 4.1,4,4,A 4.6,4.5,4.8,A 4.5,4.4,4.5,A 4.3,4.8,4.7,A 4.1,4.2,4.3,A
4,4.5,4.5,A 4.3,4.5,4.5,A 4.4,4.5,4.5,A 4.6,4.3,4.3,A 4.6,4.3,4.3,A 4.4,4.3,4.4,A 4.5,4.4,4.5,A 4.4,4.4,4.5,A
4.3,4.4,4.4,A 4.3,4.4,4.4,A 4.5,4.5,4.5,A 4.2,4.2,4.3,A 4.2,4.3,4.3,A 4.4,4.4,4.6,A 4.5,4.8,4.8,A 4.2,4.2,4.2,A
4.2,4.2,4.3,A 4.4,4.4,4.6,A 4.2,4.3,4.2,A 4.2,4.3,4.3,A 4.2,4.3,4.4,A 4.2,4.3,4.3,A
4.1,3.8,3.9,B 3.9,3.8,3.8,B 4.1,3.8,3.9,B 3.6,3.8,3.7,B 4.1,4,4,B 4.6,4,4,B 3.8,3.5,3.5,B 3.7,3.7,3.8,B 3.8,4,4,B
4.1,4.1,4.2,B 3.7,3.7,3.5,B 4,3.2,3,B 4,3,3,B 3.7,4,4,B 4,4.1,4.1,B 4,3.4,3.2,B 4,3.5,3.5,B 4,4,4,B 4,3.5,3.5,B
3.4,3.6,3.5,B 3.8,3.9,4,B 4,4,4,B 4.3,4,4,B 4,4,4,B 4.1,3.6,3.8,B 4,3.8,3.8,B 3.9,4,4,B 4,3.9,3.9,B 4,3.9,3.9,B
3.8,3.2,3.4,B 4,3.8,3.9,B 4.1,4,4,B 3.8,3.8,3.9,B 3.8,3.5,3.7,B 3.8,3.6,3.8,B 4.1,4.2,4,B 4,3.8,3.7,B 4.3,4,4,B
3.9,4,4,B 4.2,4.3,4.3,B 4,4,4,B 4.1,4,4,B 3.9,3.9,3.8,B 3.9,4,4,B
2.4,2,2,F 2.6,2.5,2.5,F 3.2,2.2,2.2,F 2.9,2,2.5,F 2.3,2,2,F 3.1,2.5,3,F 2.6,2,2.5,F 2.9,2,2,F 2.5,2,2,F 2.6,2.2,2.4,F
3.2,2,2,F 4.1,3,3,F 2.7,2,2,F 2.3,2,2.5,F 2.7,2.5,2,F 2.9,2.5,2.5,F 2.9,2.5,2.5,F
```

الملحق (9) – بيانات الاختبار بصيغة ملف ARFF حسب المنهج /2/

يتضمن بيانات الاختبار التي تم تنفيذها بشكل عملي في هذا البحث حسب المنهج /2/ المتضمن /4/ معايير مختلفة. تم ترتيب البيانات حسب انتظامها حيث ضمت بيانات الاختبار الست الأولى بيانات منتظمة أما البيانات المتبقية فهي بيانات اختبار غير منظمة.

```
@relation Evaluation5_3
@attribute Basic-Skills numeric
@attribute Service.Functions numeric
@attribute Org.Objectives numeric
@attribute Result {A,B,F}
@data
5,5,5,?
4,4,4,?
3,3,3,?
3,4,4,?
4,3.8,3.9,?
3,3.5,4,?
2,2,5,?
3,4,5,?
5,5,2,?
5,2,4,?
2,4,3,?
5,2,4,?
```

الملحق (10) – بيانات التفقد اليومي المدمجة مع المعلومات الأساسية

Level	Gender	Certificate	Service Duration	Absence Rate	Leave Rate	Rest Days
1	M	PhD	42	1	10	5
1	M	Bachelor	20	1	10	5
1	M	Bachelor	34	1	20	5
1	M	Bachelor	25	1	30	5
1	M	Bachelor	40	1	10	5
1	M	Bachelor	30	1	20	5
1	M	PhD	25	1	60	5
1	M	Bachelor	15	1	50	5
1	M	Bachelor	22	1	20	5
1	M	Bachelor	35	1	10	20
1	M	Bachelor	25	1	20	5
1	M	Bachelor	28	1	20	5
1	M	Bachelor	25	1	30	5
1	M	Bachelor	35	1	10	5
1	M	PhD	28	1	10	5
1	M	Bachelor	35	2	30	20
1	M	Bachelor	20	1	50	10
1	M	Bachelor	27	1	50	10
1	F	Bachelor	27	1	50	20
1	M	Bachelor	6	1	20	5
1	F	Bachelor	13	1	30	10
1	F	Bachelor	17	1	20	30
1	M	Bachelor	17	1	30	10
2	M	HighSchool	24	1	30	10
1	M	Bachelor	33	1	30	10
1	M	Bachelor	30	1	50	5
1	M	Bachelor	24	2	40	10
1	M	Bachelor	33	1	40	5
1	F	Bachelor	20	2	50	20
2	F	Institute	21	1	40	20
3	M	Educated	20	4	80	20
2	M	Institute	25	1	40	10
1	M	Bachelor	18	1	30	10
2	M	Institute	30	5	80	20
2	M	HighSchool	10	1	40	10
3	M	Preparatory	25	3	40	10
2	M	HighSchool	13	1	50	10
1	M	Bachelor	5	2	80	10
1	M	Bachelor	27	1	90	5
1	M	Bachelor	17	5	100	20
1	M	Bachelor	4	4	100	20
1	M	Bachelor	9	1	70	10
2	F	Institute	5	4	100	20
1	M	Bachelor	7	3	60	5
1	M	Bachelor	13	2	70	10
4	M	Educated	3	5	100	30
4	M	Educated	3	2	100	5
3	M	Preparatory	15	3	100	20
4	M	Educated	3	1	80	5
4	M	Educated	14	1	60	5
4	M	Educated	9	5	100	30
3	M	Preparatory	20	4	100	15
4	M	Educated	20	2	100	10
2	M	Institute	12	1	80	5

2	M	Institute	4	2	100	10
2	M	Institute	4	2	100	10
2	M	Institute	4	2	100	10
2	M	Institute	4	5	100	10
4	M	Educated	30	1	20	5
4	M	Educated	21	1	20	5
2	F	Institute	4	1	70	10
3	F	Preparatory	13	1	70	10
2	M	HighSchool	30	1	70	15
4	M	Educated	11	2	90	10
2	M	Institute	5	1	80	5
2	M	HighSchool	5	1	80	5
1	M	Bachelor	3	4	100	15
2	M	Institute	25	2	80	10
2	M	HighSchool	5	1	80	5
2	M	HighSchool	5	1	70	5
1	F	Bachelor	5	2	100	10
4	M	Educated	22	2	90	10
4	M	Educated	17	1	20	5
1	F	Bacholer	28	5	100	20
4	M	Educated	22	5	40	5
2	M	HighSchool	7	3	70	5
2	M	Institute	5	1	80	5
2	M	Institute	5	5	100	20
2	F	Institute	25	1	100	10
2	M	Institute	4	5	100	20
2	M	HighSchool	5	1	100	20
1	M	Bachelor	10	1	60	5
4	F	Educated	30	2	80	20
1	M	Bacholer	32	2	50	15
1	F	Bacholer	17	1	90	30
4	M	Educated	28	2	20	5
2	M	HighSchool	19	5	80	5
2	F	Institute	15	5	100	30
2	M	HighSchool	27	2	60	5
2	M	HighSchool	10	2	70	5
2	M	Institute	4	1	100	5
2	M	HighSchool	4	3	100	5
4	F	Educated	15	5	100	30
4	M	Educated	15	2	70	5
1	M	Bachelor	35	4	100	5
1	M	Bachelor	28	3	100	20
2	M	HighSchool	22	5	100	20
2	F	HighSchool	21	4	100	20
1	M	Bachelor	10	4	100	15
2	F	Institute	7	4	100	20

Abstract

Introduction:

Valuable knowledge implicit inside HR data which can be discovered through data mining methods. The success of any organization essentially depends on this knowledge that is used as raw material in decision making processes. This research presents a study for implementation of data mining approach in HRM field through analysis of personnel data which collected from some S.A.R organizations. Several tools have been used such as Classifying, Clustering and Prediction. Data related to personnel's performance appraisal and data accomplished by personnel's attendance recording have been used in this study to achieve many objectives such as predicting suitable job function of employee and clustering personnel in meaningful groups according to their attendance. Finally, this study supports several studies confirm the importance of using DM in HRM field and results that DM can be hopeful tool for dealing HR activities.

The Research Aims:

Syrian organizations keep large amounts of information and data about their personnel in their IT systems. This information, however, is often left unutilized or may be analyzed through statistical methods. In this study, DM is considered a solution for analyzing HR data and explore knowledge from data stored in some Syrian organization through two major stages:

Stage A: Using results of Semi-Annual performance evaluation process to build prototype showed in (Fig. 6) to accomplish two tasks:

1. Building a models to predict appropriate job function for an employee through majority principle and using high accuracy result to increase the number of training data and make it self-learning model.
2. Choose most important attributes that used in classify methods to use it in personnel selection and recruitment.

Stage B: Using data of Time & Attendance to analysis personnel activity through clustering methods and building many meaningful groups.

The methodology:

Analytical applied methodology is used in this study in order to:

1. analyses the methods of collecting data related to Semi-Annual performance process then suggestion suitable form to collect this data.
2. analyses the methods of collecting data related to Time & Attendance process then suggestion a suitable mode which contains required data.
3. Analysis the methods of data mining to choose the suitable tools that can be used in this study.
4. Finally, building prototype showed in (Fig. 6) to accomplish two tasks explored before.

The Results:

This study emphasize the importance of using DM as AI tool to retrieve information and knowledge from HR data stored in IT systems. DM techniques such as Classification, Prediction and Clustering was used to discover knowledge from many data sets include results of evaluation process and Time & Attendance records. But this study has its limitations especially lack of data collected because of limited access to HR records. Consequently, for further understanding it is important to implement this methodology on huge database that has many subsets of HR activities and using advanced DM technique such as fuzzy DM to find more rules and relation between data.

Finally, The described methodology is not useful only in HRM but in all sectors of enterprise by the explanation of data collected from different departments of organization and discovering rules that make faster and better decisions.

Keywords: Artificial Intelligence, Data Mining, Human Recourse Management.