

[نموذج تنبؤي لترشيح المتقدمين للبرامج التدريبية بأمانة العاصمة المقدسة من خلال اكتشاف المعرفة بتنقيب البيانات]

إعداد الباحثان:

[رائد بن محمد خان]

[طالب دكتوراه / إدارة المعرفة/ المملكة العربية السعودية/ جامعة الملك عبد العزيز /قسم علم المعلومات]

[د. عبد الرحمن بن عبيد القرني]

[أستاذ مشارك / إدارة المعرفة]

المستخلص:

هدفت هذه الدراسة إلى تطبيق تجربة عملية لتقديم نموذج تنبؤي لعملية ترشيح المتقدمين للبرامج التدريبية التي تقدمها الإدارة العامة للموارد البشرية في أمانة العاصمة المقدسة بالتعاون مع الجامعات و المعاهد المعتمدة محلياً و دولياً , و استعانت الدراسة بالمنهج التجريبي من خلال التجربة العملية و ذلك بتطبيق خوارزميات التنقيب في البيانات على عينة من بيانات الموظفين المتقدمين على البرامج التدريبية المقدمة من الإدارة العامة للموارد البشرية في الأمانة للعام التدريبي 1442هـ حيث اشتملت العينة على (1657) متقدم , حيث قدمت التجربة التطبيقية نموذج تنبؤي لدعم متخذي القرار لقبول الموظفين المرشحين للبرامج التدريبية التي تقدمها الإدارة العامة للموارد البشرية , حيث يعتبر التدريب أحد أهم أنشطة تنمية العنصر البشري بأمانة العاصمة المقدسة , و من أبرز نتائج التجربة التطبيقية أن تطبيق اكتشاف المعرفة من خلال تنقيب البيانات يساهم في التحول الرقمي و هو من ضمن الركائز الأساسية لتحقيق رؤية المملكة 2030 , كما يؤثر تطبيق اكتشاف المعرفة من خلال تنقيب البيانات على فعالية القرارات المتخذة لتحقيق أهداف أمانة العاصمة المقدسة و التي تحقق رؤية الأمانة , كما يساهم تطبيق اكتشاف المعرفة من خلال تنقيب البيانات في تقليص دورة اتخاذ القرار و بالتالي اتخاذ القرارات في وقت أسرع و بجهد أقل , كما تؤثر النتائج المتحصلة عليها من اكتشاف المعرفة بتنقيب البيانات على توجهات جميع المستويات الإدارية الاستراتيجية و التنفيذية و التشغيلية و ذلك بالاستفادة من كل ما يساهم في الرقابة و التصحيح للانحرافات عن تحقيق الأهداف و ذلك عن طريق مخرجات تقارير الأنظمة التي تعتمد على تقنيات التنقيب في البيانات , و أوصت الدراسة بعدة توصيات من أهمها تبني الإدارة العليا في الأمانة لتوجهات التحول الرقمي و التوعية أو التحفيز للمستويات الأدنى و الأفراد لهذا التوجه الوطني , و أوصت بتصميم و تنفيذ نظام آلي محوسب للترشيح على البرامج التدريبية التي تطرحها أمانة العاصمة المقدسة لتسهيل إجراءات الترشيح على البرامج التدريبية , و ضمان دقة البيانات , و إنشاء مستودع للبيانات الخاصة بجميع بيانات الأمانة , الذي يفترض به أن يحتوي على عدة متاجر للبيانات لمختلف الأنشطة في الأمانة مثل المشاريع و الخدمات و الطرق و غيرها , و إنشاء قاعدة بيانات مخصصة لبيانات الموارد البشرية على وجه الخصوص.

الكلمات المفتاحية: اكتشاف المعرفة , تنقيب البيانات , أمانة العاصمة المقدسة.

[A Predictive Model for Nominating Applicants for Training Programs in Holy Makkah Municipality by Discovering Knowledge through Data Mining]

Abstract:

This study aimed to apply practical experience to provide a predictive model for the process of nominating applicants for training programs provided by general administration of human resources in Holy Makkah Municipality by cooperation with universities and institutes accredited locally and internationally. The study used the experimental method through practical experience by applying data mining algorithms in a sample of applicant employee data on training programs provided by HR in municipality for training year 1442\2020, where the sample included (1657) applicants, where the applied experience provided a predictive model to support decision-makers to accept employees nominated for training programs provided by HR , Where training is considered one of the most important activities for developing human element in the holy Makkah municipality, and one of the most important results of the applied experiment is that the application of knowledge discovery through data mining contributes to digital transformation and is among the main pillars of achieving the Kingdom's vision 2030, and the application of knowledge discovery also affects Data exploration on the effectiveness of the decisions taken to achieve the objectives and visions of the Holy Makkah municipality, and the application of knowledge discovery through data mining contributes to reducing the decision-making cycle and thus making decisions in a faster time and with less effort. The results obtained from knowledge discovery by data mining also affect the directions of all strategic, executive, and operational administrative levels by making use of everything that contributes to monitoring and correcting deviations from achieving the goals through the outputs of the reports of systems that rely on data mining techniques, and the study recommended several recommendations, the most important of which is the adoption of the senior management in municipality of the directions of digital transformation and awareness or motivation of levels . Also it recommended the design and implementation of an automated computerized system for the nomination of the training programs offered by the Holy Makkah Municipality to facilitate the nomination procedures for training programs, ensure the accuracy of the data, and establish a data warehouse for all the municipality's data, which is supposed to contain several Data Marts for various activities in the municipality, such as projects, services, roads, etc., and the creation of a database dedicated to human resources data in particular.

Keywords: Knowledge discovery, Data mining, Holy Makkah Municipality.

المقدمة:

تسعى المنظمات و الكيانات إلى الحصول على المعرفة بشتى صورها وأشكالها من مصادرها المتعددة نظراً لأهمية المعرفة في إتخاذ القرارات من الناحية الإدارية أو المالية وزيادة الكفاءة بشكل عام. حيث تتعدد المصادر للمعرفة كالوثائق و اللوائح و التقارير و قواعد البيانات , هذا في حالة المعرفة الصريحة . أما في حالة المعرفة الضمنية فهناك رأس المال الفكري البشري مثل الخبرات الوظيفية المتراكمة و براءات الاختراعات و التي تستدعي التعامل مع العنصر البشري و دراسة سلوكه بالإضافة إلى الإعتماد على الأنظمة الحاسوبية في تحليل و تخزين هذه المعرفة (الدوري و أحمد 2007). و من الأمثلة على المعرفة الضمنية ما يتم إكتشافه من الأنماط عبر دراسة سلوك معين لكيان ما مثل مجموعة طلاب, أو جزء معين من قواعد البيانات لدى منظمة معينة, أو منحنى معين لاستخدام نظام ما , حيث ينتج عن دراسة و تحليل هذا المنحنى معرفة يتم إكتشافها و الحصول عليها بطرق علمية و إحصائية محوسبة معينة تساعد في إتخاذ قرار بناءً على المعرفة المكتشفة (فرنانديز و سايروال 2014). ونظراً للزيادة المضطردة في البيانات و المعلومات الناتجة عن الإنتشار الواسع للأنظمة و صفحات الويب و التي يتم تخزينها في قواعد بيانات أصبحت مع مرور الزمن بيانات "ضخمة", و أصبح من الضروري الإستفادة منها على النحو الذي يمكّن المنظمات و الشركات من زيادة كفاءتها و نمو أرباحها (الطيب و الرباعي 2019).

و يعتبر إكتشاف المعرفة في قواعد البيانات من خلال تنقيب البيانات مستمداً من العديد من المجالات بما في ذلك النظم الخبيرة ، والتعلم الآلي ، وقواعد البيانات الذكية ، واكتساب المعرفة ، والاستدلال القائم على الحالات والإحصاءات (Piatetsky-Shapiro 1989).

الجدير بالذكر أن إكتشاف المعرفة من خلال تنقيب البيانات يعتبر عملية مرتبطة بتحليل البيانات من وجهات نظر مختلفة ودمجها في معلومات مفيدة، ومن ثم يمكن استخدام المعلومات في زيادة الإيرادات والتقليل من التكاليف وزيادة الحصة السوقية وتحسين إدارة العمليات و جدولة المشاريع و التخطيط في المنظمات, و بالتالي تتم عملية إتخاذ القرارات على أساس أفضل المعلومات التي يتشاركها الفرد مع الآخرين، فالأفراد عادة بحاجة إلى الدعم عند إتخاذ القرارات ويتم ذلك من خلال تقديم المعلومات حول الخيارات والمخاطر، والمعلومات المطلوبة تختلف باختلاف نوع الاستخدام فبعض القرارات تكون من البساطة بحيث تكفي معلومات بسيطة لأخذ القرار، لكن عندما يتعلق الأمر بقرارات حاسمة فهنا تختلف طبيعة المعلومات المستخدمة لإتخاذ القرار (سيد 2013).

مشكلة الدراسة :

تمتلك المنظمات الكثير من البيانات بكمية ضخمة وغير منظمة في الأغلب، وهو ما يجعلها تفقد الكثير من العوائد من هذه البيانات إلى أنها غير مصنفة وليست منظمة ولا يوجد لدى تلك الجهات برمجيات متخصصة في إدارة محتوى تلك البيانات الضخمة وتحليلها، ولا يوجد محرك بحث قادر على البحث والاسترجاع الفعال في هذا الكم الهائل من البيانات والمعلومات، ولا شك أن متخذ القرار يحتاج بانتظام إلى توفر نتائج تحليل البيانات الضخمة لتكون ملائمة له وتناسب حاجاته عند اتخاذ القرار (الأكلي 2019).

و بناءً عليه أوصت العديد من الدراسات بضرورة استخدام تكنولوجيا التنقيب عن البيانات بمفاهيمها المعرفية والتقنية مع ما يلزم لذلك من أساليب وأدوات لتعزيز ثقافة المؤسسة المعرفية، وتوفير كل ما يلزم من مخازن ومستودعات البيانات في تكوين معارف جديدة من شأنها أن تساعد في تطوير أداء المؤسسة، فلذلك أوصت دراسة (بلعباس و آخرون 2017) بالاهتمام بموضوع تبني المفاهيم الحديثة في نجاح نظم دعم القرار في المؤسسة، ولغايات توفير المعلومات المناسبة في الوقت المناسب، للمساعدة في اتخاذ القرارات السليمة، و ضرورة عقد العديد من الندوات والمؤتمرات فيما يتعلق بالتنقيب عن البيانات، مع توضيح أوجه الاستفادة طبقاً للتطورات التكنولوجية الحادثة في المجال، والتركيز على تدريب وتطوير متخذي القرار على مثل هذه التقنيات الحديثة لكي يسهل عليهم استخدام أو تطبيق البرنامج والتوصل إلى المعلومات المطلوبة في اتخاذ قراراتهم، كما أوصت دراسة (Lawal et al. 2015) بأن إطار تنقيب البيانات وإدارة المعرفة يجب على جميع المنظمات الاستفادة منه إلى الحد الأقصى لرفع الميزة التنافسية، وكذلك أدوات التنقيب عن البيانات من أهم المحفزات التي يمكن للشركات استخدامها لتحقيق أهدافها المتمثلة في الربحية والميزة التنافسية وبناء علاقات عملاء واستراتيجيات لتنمية العنصر البشري فعالة.

و بناءً عليه تمثلت مشكلة الدراسة في التساؤل الرئيسي التالي : ما هو النموذج تنبؤي لترشيح المتقدمين للبرامج التدريبية بأمانة العاصمة المقدسة من خلال اكتشاف المعرفة بتنقيب البيانات ؟

أهمية الدراسة :

عملية اكتشاف المعرفة من خلال تنقيب البيانات تقدم تحليل للبيانات من منظورات مختلفة واستخلاص علاقات بينها وتلخيصها إلى معلومات مفيدة، مثل معلومات يمكن أن تسهم في زيادة الربحية، أو تخفيض التكاليف، أو لحل مشكلة معينة من منظور المستفيد، وكذلك اتخاذ القرار الذي يعتبر من الركائز الأساسية للمؤسسات، فهو يلعب دوراً أساسياً في تحقيق أهداف واستراتيجيات المؤسسة، كما يعتبر القرار الإستراتيجي من القرارات التي تعتمد عليها كثير من المؤسسات المهنية، وذلك لأنها تنصب بشكل أساسي في تحقيق أهداف المؤسسة، وكذلك تستمد الدراسة أهميتها من العنصر البشري وتنميته وتطويره فهو يمثل اليوم كأحد أهم الوظائف الإدارية التي لا تقل أهمية عن باقي الوظائف الأخرى كالإنتاج والعقود والتسويق وذلك لما للعنصر البشري من تأثيره على الكفاءة الإنتاجية للمنشآت التي تسعى حثيثاً بغية تحقيق أهدافها من أرباح مالية ورفع جودة المنتج والخدمات المرتبطة وصولاً إلى دخول عالم الإبداع والتميز.

كما تستمد الدراسة أهميتها التطبيقية ; من المنظمة التي ستطبق عليها الدراسة , حيث تمتلك أمانة العاصمة المقدسة مركز بيانات يحتوي على بيانات ضخمة بناءً على الأنظمة المستخدمة في الأمانة لخدمة المستفيدين (مثل الموارد البشرية و الرخص التجارية و الإنشائية و البلاغات و غيرها) منذ 20 عام تقريباً بالإضافة لقاعدة بيانات مكانية (SDI) تعتمد عليها الأنظمة الإنشائية و الهندسية , و باعتبار أن البنى التحتية التقنية و ما تحتويه من بيانات مخزنة في قواعد البيانات مصدر أساسي من مصادر المعرفة , حيث بادرت الإدارة العامة لتقنية المعلومات بأمانة العاصمة المقدسة إلى تبني منهجيات و ممارسات مثل مؤشرات قياس الأداء (KPI's) بالاعتماد على أدوات و برمجيات التحليل المختلفة لهذه البيانات و تقديم نتائجها لحظياً لمتخذ القرار في الإدارة العليا لاستخدام هذه المعرفة.

أهداف الدراسة:

1. التعرف على الأساس الذي يقوم عليه اكتشاف المعرفة بتنقيب البيانات في أمانة العاصمة المقدسة.
2. التعرف على المراحل الأساسية لتصميم النموذج التنبؤي لترشيح المتقدمين على البرامج التدريبية المقدمة من الإدارة العامة للموارد البشرية في أمانة العاصمة المقدسة .
3. التعرف على المعايير العلمية المناسبة للتحقق من دقة النموذج التنبؤي المقترح.

تساؤلات الدراسة :

تحاول الدراسة الإجابة على السؤال الرئيسي من خلال التساؤلات التالية :

1. ما هو الأساس الذي يقوم عليه اكتشاف المعرفة بتنقيب البيانات في أمانة العاصمة المقدسة.
4. ما هي المراحل الأساسية لتصميم النموذج التنبؤي لترشيح المتقدمين على البرامج التدريبية المقدمة من الإدارة العامة للموارد البشرية في أمانة العاصمة المقدسة .
2. ماهي المعايير العلمية المناسبة للتحقق من دقة النموذج التنبؤي المقترح.

منهج الدراسة:

استعانت الدراسة بالمنهج التجريبي و ذلك من خلال التجربة العملية لتطبيق اكتشاف المعرفة من خلال اتباع أسلوب التصنيف بخوارزميات التنقيب في البيانات على عينة من بيانات الموظفين المتقدمين على البرامج التدريبية المقدمة من الإدارة العامة للموارد البشرية بأمانة العاصمة المقدسة , و ملاحظة دقة البيانات قبل تطبيق الخوارزميات و بعد تطبيقها و ذلك للوصول إلى أهداف الدراسة و التي من أهمها إقترح النموذج التنبؤي لترشيح المتقدمين للبرامج التدريبية بأمانة العاصمة المقدسة من خلال اكتشاف المعرفة بتنقيب البيانات.

الدراسات السابقة :

دراسة (أبو قرون 2018) بعنوان " مقارنة خوارزميات تنقيب البيانات التعليمية لدعم القرار في مؤسسات التعليم العالي السودانية حالة دراسية (أداء الطلاب بكلية علوم الحاسوب وتقانة المعلومات – جامعة الزعيم الأزهرى) " , و تناولت الدراسة استخدام تقنيات التنقيب في البيانات التعليمية التي تستخدم لدعم اتخاذ القرار في مؤسسات التعليم العالي السودانية باستخدام تقنيات (التجميع , قواعد الارتباط وأشجار

القرارات) , و توصلت الدراسة لعدة نتائج من أهمها تقديم طريقة توضح كيفية الاستفادة من خوارزميات تنقيب البيانات واستخدامها لدعم القرار في مؤسسات التعليم العالي , كما توصلت إلى الخروج بنتائج تساعد في اتخاذ القرار في جانب التنبؤ بالأداء الأكاديمي للطلاب , وأوصت بعدة توصيات من أهمها معالجة الخلل في بعض الخوارزميات الموضح بنتائج المقارنة في الفصول السابقة , توفير البيانات بصورة جيدة وخالية من مشاكل جودة البيانات , وأوصت كذلك ببناء مستودعات بيانات في جميع مؤسسات التعليم العالي السودانية للاستفادة منها في تحليل البيانات بدقة للمساهمة في عمليات صنع القرارات.

دراسة (سيد 2017) بعنوان " أدوات التنقيب عن البيانات مفتوحة المصدر دراسة تحليلية تقييمية " , و تناولت الدراسة موضوع أدوات التنقيب عن البيانات مفتوحة المصدر , حيث هدفت الدراسة إلى توضيح ماهية التنقيب عن البيانات ومهامه وتطبيقاته، وتحليل أدوات التنقيب عن البيانات مفتوحة المصدر وتقييمها، ومن ثم عقد مقارنة بين أدوات التنقيب عن البيانات مفتوحة المصدر , حيث انتهجت الدراسة المنهج التحليلي لدراسة أدوات التنقيب في البيانات مفتوحة المصدر المختلفة, حيث توصلت الدراسة إلى العديد من النتائج من أهمها أن هناك العديد من الدراسات التي تقترح معايير لتقييم أدوات التنقيب عن البيانات سواء التجارية أو مفتوحة المصدر أو المجانية , و أبرز ما أوصت به الدراسة عقد العديد من الندوات والمؤتمرات فيما يتعلق بالتنقيب عن البيانات في مجال علوم المكتبات والمعلومات، مع توضيح أوجه الاستفادة طبقاً للتطورات التكنولوجية الحادثة في المجال, و أيضاً ضرورة استخدام التنقيب عن البيانات في العديد من المجالات لسهولة تنظيم المعلومات والوصول إلى المعرفة بسهولة في ظل التضخم الهائل في البيانات.

دراسة (بلعباس و آخرون 2017) بعنوان " التنقيب في البيانات واكتشاف المعرفة كآلية لتحسين جودة القرار في المؤسسة الاقتصادية " , و تناولت الدراسة موضوع التنقيب في البيانات واكتشاف المعرفة كآلية لتحسين جودة القرار في المؤسسة الاقتصادية وهدفت الدراسة إلى عرض مفهوم وأهمية التنقيب في البيانات من أجل اكتشاف المعرفة، وذلك بتحويل البيانات من مجرد معلومات متراكمة وغير مفهومة إلى معلومات قيمة يمكن استغلالها والاستفادة منها في تحسين جودة القرار, كما هدفت بيان مهام تنقيب البيانات وآلياته ومزاياه وعناصره ومجالاته, وبينت الدراسة مفاهيم المعرفة ومشاركتها واكتشافها, كما أوصت الدراسة بعدة توصيات من أهمها: ضرورة استخدام تكنولوجيا التنقيب عن البيانات بمفاهيمها المعرفية والتقنية مع ما يلزم لذلك من أساليب وأدوات لتعزيز ثقافة المؤسسة المعرفية , وتوفير كل ما يلزم من مخازن ومستودعات البيانات في تكوين معارف جديدة من شأنها أن تساعد في تطوير أداء المؤسسة, وأوصت كذلك الاهتمام بموضوع تبني المفاهيم الحديثة في نجاح نظم دعم القرار في المؤسسة، ولغايات توفير المعلومات المناسبة في الوقت المناسب.

دراسة (الكردى و حنا 2017) بعنوان " تحسين اكتشاف المعرفة باستخدام تطبيقات التنقيب عن البيانات " , و تناولت الدراسة تحسين اكتشاف المعرفة باستخدام تطبيقات التنقيب عن البيانات حيث هدفت الدراسة إلى التطبيق العملي لآليات اكتشاف المعرفة بالتنقيب في البيانات على موقع ويب مختص بتجارة القطع الإلكترونية ومعدات الحواسيب , وهدفت الدراسة لتوضيح بعض الآليات المستخدمة في اكتشاف المعرفة

بتنقيب البيانات مثل : استخدام تقنية الترابط (حاسوبياً باستخدام برامج إدارة قواعد البيانات) لجمع عدد من جداول قاعدة البيانات الخاصة بالمنتجات و جداول المبيعات و تهدف إلى إظهار المنتجات التي تم شراؤها معاً, وكذلك إيجاد روابط بين البيانات وجعلها ذات ارتباطات تقضي إلى تحقيق استنتاجات, وأوصت الدراسة بتوظيف البحث وخدمات التنقيب عن البيانات في الشركات والمؤسسات الانتاجية الحكومية, كما أوصت باعتماد آليات اكتشاف المعرفة و تنقيب البيانات لما تقدمه من مميزات تساعد المنظمات على توقع أنشطتها مثل التخزين و التوزيع في فترة زمنية مستقبلية.

دراسة (موسى و عبد الحسن 2015) بعنوان "أثر تقنية تنقيب البيانات في تطوير عملية صنع القرارات دراسة استطلاعية في وزارة التعليم العالي والبحث العلمي العراقية", و تناولت الدراسة موضوع الدور التأثيري لاستخدام تقنية تنقيب البيانات في تطوير عملية صنع القرار في وزارة التعليم العالي و البحث العلمي العراقية, حيث هدفت الدراسة إلى المساهمة في رفع مستوى قدرة وزارة التعليم العالي و البحث العلمي العراقية على اتخاذ القرارات الرشيدة من خلال استخدام التقنيات الحديثة التي تمكنها من اكتساب المزيد من المعرفة , كما هدفت إلى التعرف على عملية صنع القرارات و مدى تأثير تقنية تنقيب البيانات في فاعليتها لدى متخذي القرارات في وزارة التعليم العالي و البحث العلمي العراقية , حيث انتهجت الدراسة المنهج المسحي لمعرفة أثر تقنية تنقيب البيانات على عملية صنع القرار , حيث توصلت الدراسة إلى العديد من النتائج من أهمها أن حاجة المنظمات إلى تقنية تنقيب البيانات تزداد كلما كان هناك كم كبير من البيانات ناتج من طبيعة حجم النشاط الذي تمارسه تلك المنظمات , و أن الاستخدامات التقليدية لتكنولوجيا المعلومات أصبحت اليوم غير كفؤة في جعل المنظمات منافسة في ظل الحاجة إلى رفع قيمة المعلومات لتوفير قدرات لإسناد عملية صنع القرار.

دراسة (الديوهجي و عبدالمحسن 2013) بعنوان "اكتشاف المعرفة و مشاركتها باستخدام لوحات عدادات الأعمال في كلية الإدارة والاقتصاد بجامعة الموصل" , و تناولت اكتشاف المعرفة و مشاركتها بكفاءة , حيث أشارت الدراسة إلى أهمية المعلومات و المعرفة التي توفرها لوحات عدادات الأعمال المقترحة لتعزيز نقاط القوة و الحد من نقاط الضعف عبر مشاركة المعرفة بين المستفيدين عن طريق تطوير لوحة العدادات المستندة على نظم ذكاء الأعمال و استخدمت الدراسة منهجية التطوير السريع للتطبيقات حيث توصلت الدراسة بأن استخدام تقنيات عدادات الأعمال يقدم تحليلات قيّمة و ذات جودة لكونها مشتقة من بيانات قد تم تهيئتها وفقاً لمتطلبات المستفيد و توصلت أيضاً أن لوحات عدادات الأعمال ساعدت في كفاءة و فاعلية أداء الكلية من خلال إيصال المعرفة في الوقت المناسب من خلال ميزات التحديث التلقائي و أوصت الدراسة بزيادة استخدام تقنيات و أساليب متقدمة في تطوير عدادات الأعمال كتلك المستندة على الابداع أو المستندة على متاجر البيانات و أوصت أيضاً على تدريب المستفيدين النهائيين على ثقافة القياس للوصول إلى المستويات المرغوبة من الأداء.

دراسة (Ristoski and Paulheim 2016) بعنوان " Semantic Web in data mining and knowledge discovery: A comprehensive survey "

و تناولت الدراسة الويب الدلالي في استخراج البيانات واكتشاف المعرفة , حيث هدفت الدراسة إلى تقديم مسح شامل للأساليب التي تجمع بين اكتشاف المعرفة وتنقيب البيانات و الويب الدلالي , و هدفت أيضاً لبيان العلاقة بين الأساليب المستندة إلى الويب الدلالية , وتقنيات الويب الدلالية , والبيانات المفتوحة المرتبطة لدعم عملية اكتشاف المعرفة, وكذلك ركزت الدراسة على أساليب اكتشاف المعرفة بتنقيب البيانات في الويب الدلالي, و توصلت الدراسة إلى العديد من النتائج من أهمها : أنه يمكن أن تساعد بيانات الويب الدلالي في تفسير الأنماط الموجودة في قواعد البيانات للمساعدة في عمليات اكتشاف المعرفة , لا سيما بالنسبة للتنقيب بالأساليب الوصفية , وتوصلت أيضاً بأنه هناك الكثير من الأعمال البحثية في مجال اكتشاف المعرفة وتنقيب البيانات , وكذلك التطبيقات موجودة في العديد من المجالات مثل التقنيات الطبية و الحيوية و التجارية , و توصلت أيضاً بأنه لا تزال هناك بعض المجالات التي لم يتم تغطيتها بحثياً بالشكل المناسب خاصة في مجال اكتشاف المعرفة في الويب الدلالي .

دراسة (Hamad & QADER 2014) بعنوان " Data Pre-processing for knowledge discovery "

و تناولت الدراسة اكتشاف المعرفة من منظور تقني متخصص حيث هدفت الدراسة لبيان أهمية تهيئة و معالجة البيانات كخطوة أساسية من خطوات اكتشاف المعرفة و اعتبرت مرحلة ما قبل المعالجة خطوة مهمة لعملية اكتشاف المعرفة ولها تأثير كبير على الدقة التنبؤية كما هدفت لبيان نتيجة استخدام طرق متعددة للمعالجة المسبقة هما (الاتساق ، التخفيض) اعتماداً على مستودع بيانات يحتوي على سمات غير متسقة ويحتوي أيضاً على سجلات مكررة و من أهم ما توصلت إليه الدراسة بأن المعالجة المسبقة للبيانات مرحلة مهمة وضرورية لاكتشاف المعرفة لأن المعالجة المسبقة للبيانات تؤدي إلى إنشاء مستودع بيانات يكون جاهزاً ومستعداً لاكتشاف المعرفة باستخدام تقنية تنقيب البيانات أو أي طريقة تحليل تؤثر على جودة ودقة تفكيك المعرفة كما توصلت إلى أن تطبيق الخوارزميات الرياضية مثل خوارزمية التناقص و خوارزمية التكرار على البيانات أدى إلى إنشاء مستودع بيانات نظيف يحتوي على سمات متسقة وخالية من السجلات والسمات المكررة خلال وقت قصير جداً وتوفير بيانات عالية الجودة لاستخراج البيانات واكتشاف المعرفة.

دراسة (Fayyad et al. 1996) بعنوان " From Data Mining to Knowledge Discovery in Databases "

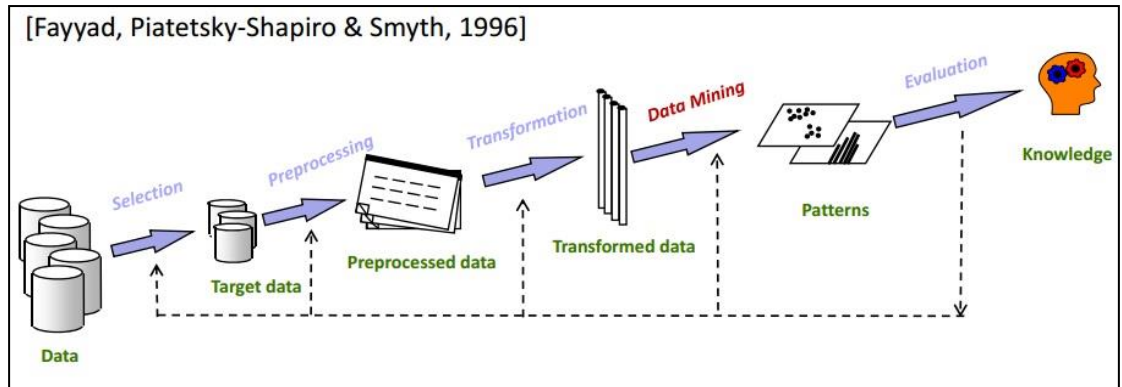
"من تنقيب البيانات إلى اكتشاف المعرفة في قواعد البيانات" , و قد تناولت الدراسة مفاهيم اكتشاف المعرفة وتنقيب البيانات والعلاقة بين اكتشاف المعرفة وتنقيب البيانات , وعلاقتها معاً ببعض المجالات الأخرى مثل تعلم الآلة والإحصاء وقواعد البيانات , وذكرت الدراسة التطبيقات الواقعية لاكتشاف المعرفة , كما ناقشت الدراسة التحديات والتوجهات البحثية في هذا المجال, كما هدفت الدراسة لبيان أهمية مساعدة البشر في استخراج المعلومات ذات الأهمية "المعرفة" من أحجام البيانات الرقمية المتنامية بسرعة هائلة, من خلال نظريات وأدوات اكتشاف المعرفة في قواعد البيانات وهو ما عبرت عنه الدراسة بمصطلح (KDD) اختصاراً للتعبير "Knowledge Discovery in Databases", كما هدفت الدراسة لبيان علاقة تنقيب البيانات باكتشاف المعرفة حيث أفادت الدراسة بأن عملية اكتشاف المعرفة هي العملية الرئيسية للحصول على

المعلومات ذات المعنى و المفيدة من البيانات , بينما تنقيب البيانات هو عملية جزئية أو فرعية من عمليات اكتشاف المعرفة , حيث تهتم عملية تنقيب البيانات بتطبيق خوارزميات محددة لاستخراج الأنماط من البيانات , و أنها تعتمد على عمليات فرعية أخرى من عمليات اكتشاف المعرفة مثل إعداد البيانات و اختيار البيانات و تنظيف البيانات , و توصلت أنه إن لم يتم اتباع عمليات اكتشاف المعرفة بشكل صحيح لا تنجح عملية اكتشاف المعرفة و من ضمنها عملية تنقيب البيانات , مما يؤدي إلى اكتشاف أنماط لا معنى لها و غير صالحة, و توصلت الدراسة لبيان مهام اكتشاف المعرفة في قواعد البيانات مثل تخزين البيانات و طريقة الوصول إليها و كيفية تجزئتها لكي تعمل بكفاءة و كذلك تفسير البيانات الناتجة و عرضها و كيفية تعامل المستخدم معها , كما أوردت الدراسة بعض التحديات و المجالات المستقبلية في مجال اكتشاف المعرفة و تنقيب البيانات مثل تضخم قواعد البيانات, و التغير السريع في المعرفة و المعلومات, و فقد و تشوش البيانات, و من أهم التطبيقات المستقبلية التي تطرقت إليها الدراسة : تنقيب اللغات الطبيعية و الاستخبارات الذكية , و تمثيل المعرفة و الأنطولوجيا.

التعليق على الدراسات:

اتفقت معظم الدراسات سواء العربية و الأجنبية التي تناولت تنقيب البيانات على أهمية تنقيب البيانات نظراً لكم الهائل والمتزايد من البيانات بغرض الوصول إلى المعرفة في قواعد البيانات المختلفة من خلال التركيز على الجانب التقني العملي، كما اتفقت في سرد وتوضيح بعض آليات و أدوات التنقيب المختلفة.

لوحظ من خلال الإطلاع على الدراسات العربية و الأجنبية المتعلقة بمجال اكتشاف المعرفة و تنقيب البيانات أنها لا تخلو من الإشارة أو الإقتباس من دراسة (Fayyad et al. 1996) بالرغم من قدمها نسبياً, فقد ساهمت هذه الدراسة في توضيح المفاهيم و النماذج الأساسية في مجال اكتشاف المعرفة من قواعد البيانات, و من أهم تلك النماذج : "نموذج إجراءات اكتشاف المعرفة من قواعد البيانات " كما هو موضح أدناه :



نموذج إجراءات اكتشاف المعرفة من قواعد البيانات (Fayyad et al. 1996)

و من خلال الدراسات السابقة تبين مدى العلاقة بين اكتشاف المعرفة و تنقيب البيانات حيث يوجد تداخل ملحوظ بين مصطلحي تنقيب البيانات و اكتشاف المعرفة, خاصة و أن المختصين اختلفوا في كيفية التسمية أيهما يشمل الآخر, اكتشاف المعرفة أم تنقيب البيانات , و بطريقة تبادلية يمكننا القول إن تنقيب البيانات هو اكتشاف المعرفة من البيانات وهي عملية تحليل البيانات من منظورات مختلفة و استخلاص علاقات بينها و تلخيصها إلى معلومات مفيدة, مثل معلومات يمكن أن تسهم في زيادة الربحية, أو تخفيض التكاليف, أو لحل مشكلة معينة من منظور المستفيد.

كما أكدت الدراسات أن اكتشاف المعرفة هي عملية الكشف والعثور على معلومات ذات فائدة من خلال استعمال مجموعة من الأدوات المعقدة مثل أدوات الإحصاء والذكاء الاصطناعي والرسوم البيانية وغالباً ما تكون هذه البيانات في قواعد بيانات أو مستودعات ضخمة , و تمتاز هذه البيانات بأنها غير مهيكلة مثل الصور و الفيديو و محادثات منصات التواصل الاجتماعي.

اتفقت معظم الدراسات التي ناقشت اكتشاف المعرفة و تنقيب البيانات على أن عملية اكتشاف المعرفة هي العملية الرئيسية للحصول على المعلومات ذات المعنى و المفيدة من البيانات , بينما تنقيب البيانات هو عملية جزئية أو فرعية من عمليات اكتشاف المعرفة , حيث تهتم عملية تنقيب البيانات بتطبيق خوارزميات محددة لاستخراج الأنماط من البيانات , و أنها تعتمد على عمليات فرعية أخرى من عمليات اكتشاف المعرفة مثل إعداد البيانات و اختيار البيانات و تنظيف البيانات , و أنه إن لم يتم اتباع عمليات اكتشاف المعرفة بشكل صحيح لا تنجح عملية اكتشاف المعرفة و من ضمنها عملية تنقيب البيانات.

المحور الأول: اكتشاف المعرفة و تنقيب البيانات:

إدارة المعرفة :

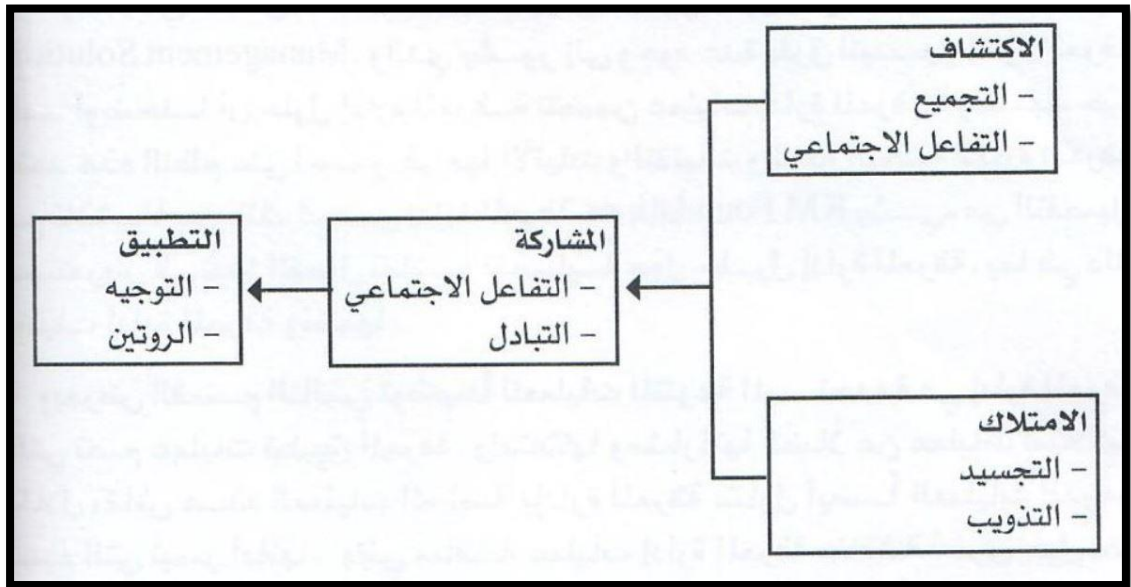
عرف (عليان 2015، 136) إدارة المعرفة بأنها " مصطلح جديد وغريب إلى حد ما , خلافا لمصطلح إدارة المعلومات ونظم المعلومات الإدارية التي تعتبر شائعة نسبياً, ورغم تزايد الاهتمام بالعقدين الأخيرين في إدارة المعرفة فإنه لا زال هناك جدلاً حول المفهوم الحقيقي لإدارة المعرفة حيث أن البعض يتصور إن إدارة المعرفة مرادف لمصطلح إدارة المعلومات, في حين يرى آخرون أنها مفهوم يتركز على الجهود الخاصة عبر الشبكات, ويرى فريق ثالث أن إدارة المعرفة هي آخر صرعات منتجي تقنية المعلومات والاستشاريين بهدف بيع حلولهم المبتكرة إلى رجال الأعمال المتلهفين لأداة يمكن أن تسهم في تحقيق الربحية والتقدم التنافسي".

كما عرف (ياسين 2007، 45) إدارة المعرفة بأنها " تسعى بشكل أساسي إلى تقديم حلول إلى قضايا و مشكلات إستقطاب و إنتاج المعرفة و المشاركة فيها بين الأفراد و الجماعات و المنظمات و داخل المنظمات في حد ذاتها ما يعني أنها تنحو إلى الإهتمام باستثمار الموارد المعرفية و بناء ذاكرة للمنظمة تركز بالأساس على تبادل المعرفة و المشاركة فيها من خلال عملية منهجية مستمرة".

عمليات إدارة المعرفة:

وفقاً لما ذهب إليه (فرنانديز و سايبروال 2014، 93) اللذين اوردا نموذج يوضح عمليات إدارة المعرفة استناداً إلى تعريف نوناكا (1994) و تعريف جرانت و جوشال (1996) , حيث تعتمد إدارة المعرفة على أربعة عمليات رئيسية هي : إكتشاف المعرفة , وإمتلاك المعرفة , و مشاركة المعرفة , و تطبيق المعرفة.

حيث تتفرع هذه العمليات لتصبح سبعة عمليات , أربعة من هذه العمليات تستند على العمليات التي اعتمدها نوناكا و هي : التفاعل الإجتماعي , والتجسيد , والتدوين , و التجميع. و الثلاث الأخرى و هي التبادل و التوجيه و الروتين تستند على رأي جرانت و جوشال (Grant and Ghoshal 1998). و فيما يلي شكل يوضح عمليات إدارة المعرفة حسب ما شرح أعلاه :



شكل (1) : عمليات إدارة المعرفة (فرنانديز و سايبروال 2014)

مفهوم اكتشاف المعرفة كعملية من عمليات إدارة المعرفة :

إشارة إلى الشكل (1) و الذي يوضح فيه الباحثان (فرنانديز و سايبروال) عملية اكتشاف المعرفة و العمليات الفرعية لها , حيث يضيف الباحثان بأنه " يمكن تعريف اكتشاف المعرفة بأنه تطوير معرفة ضمنية أو صريحة جديدة من البيانات والمعلومات أو من خلال دمج المعرفة السابقة. و يعتمد اكتشاف المعرفة الصريحة الجديدة بشكل مباشر على التجميع [combination] ، في حين يعتمد اكتشاف المعرفة الضمنية الجديدة بشكل مباشر

على التفاعل الاجتماعي [socialization]. وفي كلتا الحالتين ، يتم اكتشاف المعرفة الجديدة من خلال تجميع المعرفة من مجالين متميزين أو أكثر مع معرفة صريحة من مستمدة من مجالين تم التوفيق بينهما من خلال التجميع ، بينما المعرفة الضمنية من مجالين يتم تجميعهما من خلال التنشئة الاجتماعية".

مفهوم التجميع لاكتشاف المعرفة :

و هو من العمليات الفرعية لاكتشاف المعرفة الصريحة ، حيث يعرف عملية التجميع الباحثان (فرنانديز و سايبروال 2014، 94) نقلاً عن (Nonaka 1994) بأنها تلك العملية التي يتم من خلالها تجميع أجزاء متعددة من المعرفة الصريحة (و / أو البيانات و / أو المعلومات) لإملاك مجموعات جديدة أكثر تعقيداً من المعرفة الصريحة.

حيث يفسر الباحثان ذلك بمثال " أنه عند عرض اقتراح جديد للعميل ، يمكن دمج -تجميع- البيانات الصريحة والمعلومات والمعارف في المقترحات السابقة في الاقتراح الجديد. حيث يمكن استخدام تقنيات تنقيب البيانات للكشف عن علاقات جديدة بين البيانات الصريحة التي قد تؤدي إلى إنشاء نماذج تنبؤية أو تصنيفية لخلق معرفة جديدة".

مفهوم التفاعل الاجتماعي لاكتشاف المعرفة :

و هو من العمليات الفرعية لاكتشاف المعرفة الضمنية داخل المنظمة ، حيث يعرف الباحثان (فرنانديز و سايبروال 2014، 95) التفاعل الاجتماعي بأنه " هو حصر المعرفة الضمنية لدى الأفراد ، و عادة يتم ذلك من خلال الأنشطة المشتركة بدلاً من التعليمات المكتوبة أو الشفهية فعلى سبيل المثال ، من خلال نقل الأفكار والصور ، حيث تساعد هذه الأنشطة الموظفين الجدد على معرفة طريقة تفكير الآخرين و أساليب التعامل في المنظمة" حيث أورد الباحثان عدة ممارسات من ضمنها الاحاديث الجانبية بين الموظفين.

نشأة و تطور مصطلح اكتشاف المعرفة في الأبحاث و الدراسات العلمية :

من خلال البحث و الاستقصاء للمصطلح في الأبحاث و الدراسات العلمية في مجال اكتشاف المعرفة ، تم التوصل إلى الآتي :

- في العام 1987م قام الباحث "Piatetsky-Shapiro" بالبحث في طرق اكتشاف الأنماط تلقائياً بمجال هندسة قواعد البيانات حيث حضر ورشة العمل المهمة بمجال البيانات بعنوان "استخراج المعرفة من البيانات"، في حينها كانت أول ورشة عمل كانت ضمن مؤتمر دولي تقوم بعرض تحليل البيانات التاريخية لعدد 50000 مريض في ستانفورد و تبحث عن الآثار الجانبية للأدوية.
- في العام 1989م ظهر مصطلح "اكتشاف المعرفة أول مرة " Knowledge Discovery"، و ذلك في الورقة المقدمة من الباحث "Piatetsky-Shapiro" خلال مناقشته لورشة العمل التي أقيمت ضمن المؤتمرات الدولية المشتركة حول الذكاء الاصطناعي "IJCAI-89 International Joint Conferences on Artificial Intelligence"، و التي كانت بعنوان " Knowledge Discovery in Real Databases: A

"Report on the IJCAI-89 Workshop", حيث تمت الإشارة بأن مصطلح اكتشاف في قواعد البيانات موضوعًا مثيرًا للاهتمام ، مستمدًا من العديد من المجالات بما في ذلك النظم الخبيرة ، والتعلم الآلي ، وقواعد البيانات الذكية ، واكتساب المعرفة ، والاستدلال القائم على الحالات والإحصاءات (Piatetsky-Shapiro 1989).

- تم إيراد مصطلح اكتشاف المعرفة في العام 1994م حسبما أورد الباحثان فرنانديز و سابيروال نقلاً عن نوناكا (1994) في دراسته التي تضمنت "أن إدارة المعرفة تعتمد على أربعة عمليات رئيسية هي : إكتشاف المعرفة , وإملاك المعرفة , و مشاركة المعرفة , و تطبيق المعرفة (فرنانديز و سابيروال 2014).
- في العام 1996م , قدم الباحث (Usama Fayyad) دراسة من أهم الدراسات التي لفتت انتباه المجتمع العلمي و التقني بمجال اكتشاف المعرفة و التي كانت بعنوان " From Data Mining to Knowledge Discovery in Databases " "من تنقيب البيانات إلى اكتشاف المعرفة في قواعد البيانات" و التي كانت ضمن مؤتمر IJCAI-96, و التي ساهمت في توضيح العديد من المفاهيم من أهمها : توضيح العلاقة بين اكتشاف المعرفة و تنقيب البيانات و لمحة عامة عن عملية اكتشاف المعرفة من قواعد البيانات , وأساليب استخراج البيانات الأساسية (2019 ijcai.org).

المحور الثالث : تنقيب البيانات:

مفهوم تنقيب البيانات :

يعتبر (Avison & Shah 1997) تنقيب البيانات "أحد تكنولوجيات الذكاء الاصطناعي بالإضافة إلى الانظمة الخبيرة والشبكات العصبية ويهدف الى تمكين المنظمة من الاستغلال الأمثل لبياناتها فهي تحاول إيجاد المعلومات في مجاميع البيانات الكبيرة التي قد لا تعلم المنظمة بوجودها وكذلك إيجاد العلاقات وعمل التنبؤات".

كما عرف (Romney & Steinbart 2000) تنقيب البيانات بأنه "تقنية أو أسلوب يمكن من خلاله الوصول الى المعلومات المخزونة في مستودع ويتضمن استخدام أنواع من التحليلات منها التحليل الاحصائي لاكتشاف العلاقات الخفية في البيانات".

و يعرف (سيد 2015) تنقيب البيانات بأنه " عملية متطورة تقوم باشتقاق البيانات المطلوبة والفعالة والشاملة من كم هائل من البيانات طبقاً لأهداف مسبقة تجارية، ويعد البعض التنقيب عن البيانات مصطلحاً شائعاً لاكتشاف المعرفة، في حين يضع البعض التنقيب عن البيانات كخطوات أساسية في عملية اكتشاف المعرفة".

و يشير (Kamber & Han 2000) بأن " التنقيب عن البيانات ظهر في أواخر الثمانينات والذي كان مجالاً حديثاً مع قيمة بحثية مرتفعة في دراسة قواعد البيانات والذكاء الصناعي وتكنولوجيا قواعد البيانات وتعلم الآلة والإحصائيات وعرض البيانات وغيرها من المجالات النظرية والتقنية".

و من وجهة نظر أخرى تضيف (سيد 2013) بأن تنقيب البيانات " عملية مرتبطة بتحليل البيانات من وجهات نظر مختلفة ودمجها في معلومات مفيدة، ومن ثم يمكن استخدام المعلومات في زيادة الإيرادات

والتقليل من التكاليف أو الاثنين معا. ولقد تم تصميم برامج خاصة بالتنقيب عن البيانات تتميز باحتوائها على مجموعة من أدوات التحليل تستخدم في تحليل البيانات، حيث تمكن المستفيد من تحليل البيانات من أبعاد وزوايا مختلفة مع إمكانية تصنيفها وتلخيص العالقة بينهم".

ومن التعريفات الشاملة من وجهة نظر الباحث , ما ذكره (Kantardzic 2003) حيث عرف تنقيب البيانات بأنه " هو البحث عن معلومات جديدة وقيمة و غير بديهية في كميات كبيرة من البيانات من خلال جهد تعاوني بين البشر و أجهزة الكمبيوتر حيث يتم تحقيق أفضل النتائج من خلال تحقيق التوازن بين المعرفة لدى الإنسان وقدرات البحث لدى أجهزة الكمبيوتر".

مفهوم تنقيب البيانات وإدارة المعرفة:

ذكر (فرنانديز 2014، 75) بأن "إدارة المعرفة تعنى بالقيام بما هو مطلوب للحصول على أقصى استفادة من موارد المعرفة كما تعتبر إدارة المعرفة أحد المتطلبات البالغة الأهمية التي تشجع على اكتشاف معرفة الشركة والتقاطها ومشاركتها وتطبيقها".

وفي هذا السياق يوضح (أبو بكر 2017) أن من أنشطة و عمليات إدارة المعرفة: عملية إكتشاف المعرفة و هي خلاصة عملية التنقيب عن البيانات حيث أن عملية التنقيب هي البحث عن المعرفة من البيانات دون فرضيات مسبقة عما يمكن أن تكون هذه المعرفة.

ولكي نربط بين اكتشاف المعرفة و عملية التنقيب في البيانات يوضح (الكردى و حنا 2017) بأن عمليات اكتشاف المعرفة تشتمل على أو تتضمن عملية التنقيب عن البيانات من خلال العمليات التالية :

- 1- اكتشاف البيانات :
وهي مرحلة جمع البيانات وتشمل كشف وتحديد وتوصيف البيانات.
- 2- تصفية البيانات :
حذف البيانات المتضاربة وغير المهمة.
- 3- تكامل البيانات :
تجميع البيانات التي قد تكون من مصادر متعددة.
- 4- اختيار البيانات :
استرجاع كل البيانات التي لها علاقة بمهمة التحميل من قاعدة البيانات.
- 5- تحويل البيانات :
تحويل البيانات إلى نماذج مخصصة للتنقيب بواسطة عمليات التجميع.
- 6- تنقيب البيانات :

استخدام طرق ذكية تطبق لاستخلاص أنماط البيانات.

7- تقييم النمط :

تميز الأنماط المهمة حيث تمثل قاعدة المعرفة ببعض المقاييس المهمة.

8- عرض أو تمثيل المعرفة :

تقنيات تمثيل المعرفة والرؤية تستخدم لتقديم المعرفة المنقبة عنها للمستخدم.

و اعتبر البعض عملية التنقيب في البيانات من العمليات المهمة ضمن مراحل تشارك المعرفة في المنظمات , حيث يشير (راتب و آخرون 2011) بأن مهام تشارك المعرفة تشتمل على الآتي :

الأولى :إنشاء الشبكة الداخلية

وتستخدم لدعم الوصول وتبادل المعرفة بالمنظمة.

الثانية :إنشاء مستودع البيانات/المعرفة

ان مستودع المعرفة يتشكل من كل المعلومات الصريحة المرمزة .

الثالثة :تنفيذ أدوات القرار

أن جعل ثروة المعرفة متاحة في جميع أنحاء الشركة يحقق منفعة حقيقة للشركة التي ترغب في تحسين قدرة العاملين على صنع القرار.

الرابعة :تحقيق التشارك الجماعي لدعم التشارك

أن التشارك الجماعي ينظر إليه على انه وسيلة لتقاسم الأفكار وطريقة للتدفق الحر للمعرفة بأكثر مما تسمح به مستودعات المعرفة أو أنظمة دعم القرار أن التشارك مستخدم بقوة للتوليد الجماعي للمعرفة ونقلها وتقاسمها وان الكثير من الشركات في سياق هذه الممارسة تستخدم التكنولوجيا لتحقيق فضاءات التشارك في أي وقت وفي أي مكان.

المجالات و المفاهيم المرتبطة بتنقيب البيانات:

يشير (مليك و صباح 2017) بأنه هناك عدة ميادين لتطبيق التنقيب في البيانات في منظمات الأعمال نذكر منها :

- التسويق:

دراسات التسويق المستهدف بما في ذلك الحصص السوقية. و قد ساعدت هذه الأساليب التسويق على استخدام نهج تخصيص الزبائن وفقا إلى الحقائق الديمغرافية (السكانية) الأساسية مثل الجنس والعمر والمجموعات وكذلك أنماطهم الشرائية .

- التجزئة:

- لقد استخدمت أساليب التنقيب في البيانات بصورة فعالة في التنبؤ بالمبيعات حيث أخذت العديد من المتغيرات في الدراسات مثل متغيرات السوق المتعددة، قدرات الزبائن المستندة على العادات المتبعة في الشراء. كما ساعدت أساليب مثل تحليل السلة الشرائية أو السلة السوقية كثيرا على إيجاد أي من المنتجات التي يمكن أن تشتري سوية من قبل الزبائن .

- البنوك:

لقد أثبتت تنبؤات الأعمال والمالية على أنها الأساليب الممتازة في تطبيقات أساليب التنقيب في البيانات. وقد استخدمت هذه الأساليب في إيجاد الأسعار المضمونة وتنبؤات السعر المستقبلية وأداء الأسهم. كما وقد حققت استخدامات مثل هذه الأساليب النجاح في تطوير أنظمة القياس الرقمية في تحديد مخاطر القروض والاحتيالات المالية .

- التأمين:

لقد استخدمت أساليب التنقيب في البيانات كذلك وبصورة واسعة في مجال شركات التأمين وخاصة في تخصيص مجاميع الزبائن بهدف تحديد أسعار البوليصة وذبذبات المطالبات المتوقعة مستقبلا، وتحديد المطالبات المزيفة.

- إدارة العمليات: حيث استخدمت الشبكات العصبية في عمليات التخطيط والجدولة وإدارة المشاريع بالإضافة إلى إدارة الجودة.

المحور الثاني : أساليب و طرق تنقيب البيانات :

ذكر (بلعباس و آخرون 2017) من خلال الدراسة التي ناقشت التنقيب في البيانات واكتشاف المعرفة كآلية لتحسين جودة القرار في المؤسسة الاقتصادية عدد من الطرق لتنقيب البيانات . من أهمها ما يلي :

1. التصنيف Classification :

يتمثل في تفسير أو التنبؤ بخاصية فرد ما من خلال خصائص أخرى.

2. الانحدار Regression :

يستخدم في التنبؤ بالقيم الجديدة بالاعتماد على القيم الموجودة .

3. تحليل السلاسل الزمنية Analysis Series Time : تتنبأ بالقيم غير المعروفة بالاعتماد على سلاسل تأخذ بالحسبان الخواص المميزة للزمن .
4. التنبؤ Prediction : يشبه التنبؤ التصنيف أو التقدير، ما عدا أن البيانات تصنف على أساس التنبؤ بسلوكها المستقبلي أو تقدير قيمتها المستقبلية .
5. التجميع العنقودي Clustering : البحث عن مجموعات متجانسة في مجتمع من الأفراد .
6. التلخيص Summarization : تقتطع كتل البيانات وتقديم وصفا عاما للمتغيرات وعلاقاتها.
7. قواعد الارتباط Rules Association : البحث عن علاقات موجودة بين عدة خصائص .
8. اكتشاف التسلسل Discovery Sequence : تستخدم لتحديد نماذج متسلسلة في البيانات.

مراحل تنقيب البيانات :

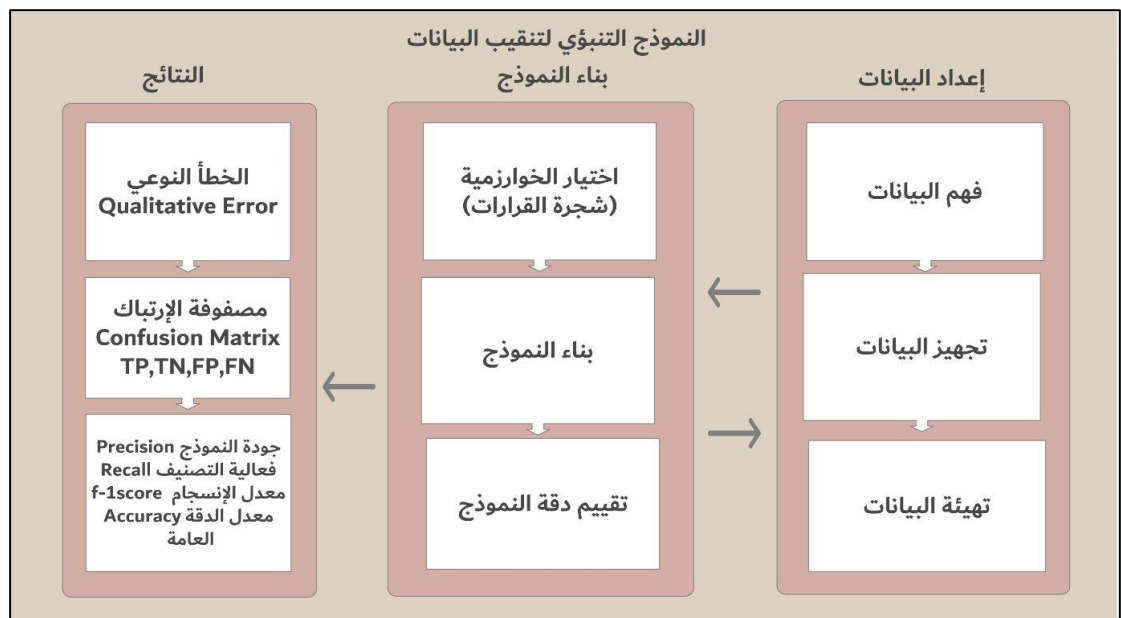
أقترح (Kamber & Han 2000) مراحل للتنقيب في البيانات لاكتشاف المعرفة , و تتلخص هذه المراحل على النحو الآتي :

- 1- تنقية البيانات: تختص هذه العملية بحذف البيانات غير المهمة والمكررة.
 - 2- تكامل البيانات: وتعمل هذه المرحلة على جمع البيانات من المصادر المختلفة.
 - 3- إختيار البيانات: وفي هذه المرحلة يتم إختيار البيانات التي سيتم عليها التحليل.
 - 4- تحويل البيانات : عملية توحيد/إدماج البيانات في أشكال محددة مناسبة حتى تكون مهيئة لعملية التنقيب.
 - 5- التنقيب عن البيانات: وتعد الخطوة الجوهرية حيث يتم استخدام التقنيات الذكية لاستنباط أنماط مفيدة من المعلومات.
 - 6- تقييم الأنماط : أي تحديد الأنماط التي تمثل المعرفة وفق المقاييس المعطاة.
 - 7- تمثيل المعرفة : وهي المرحلة الأخيرة حيث يتم اكتشاف المعرفة الجديدة، وفي هذه المرحلة يتم استخدام التقنيات المصورة Visualization لمساعدة المستخدمين على فهم وتفسير النتائج المستخرجة.
- و من خلال استعراض و مراجعة الدراسات و الإهتمامات البحثية المتعلقة بتنقيب البيانات, وجد أنه غالباً ما يرتبط مصطلح تنقيب البيانات بمصطلح البيانات الضخمة , فبالنظر إلى تعريفات مصطلح تنقيب البيانات نرى أن الباحثين أتفقوا في الإشارة بأن عمليات التنقيب تتم في مجموعه كبيرة أو هائلة أو ضخمة من البيانات , فقد عرفت عدة مؤسسات و جهات مصطلح البيانات الضخمة حسب رؤيتها و أنشطتها و من أبرز من عرف البيانات الضخمة مؤسسة غارتر للأبحاث و الاستشارات.

المحور الثالث : خطوات تصميم التجربة لاكتشاف المعرفة بتنقيب البيانات لتقديم نموذج تنبؤي لترشيح المتقدمين على البرامج التدريبية :

بعد الاطلاع على الأدبيات، والدراسات السابقة ذات الصلة بموضوع الدراسة الحالية، وفي ضوء معطيات وتساؤلات الدراسة وأهدافها تم تصميم التجربة بطريقة المجموعة الواحدة ، وفيما يلي عرض لكيفية تصميمها، والإجراءات المتبعة للتحقق من دقة النموذج التنبؤي الناتج عنها.

تم تقسيم مراحل تصميم التجربة إلى ثلاثة مراحل رئيسية تتفرع منها عدد من المهام الفرعية. الشكل التالي يوضح المراحل الأساسية لتصميم التجربة :



المراحل الأساسية لتصميم التجربة (من تصميم الباحث)

المرحلة الأولى : إعداد البيانات :

تشتمل هذه المرحلة على المهام الضرورية لجعل البيانات متاحة و مناسبة لعمليات التنقيب و تحقيق أهداف الدراسة , ولكي تتم هذه المهمة لابد من الحصول عليها من مصادرها المعتمدة من متخذ القرار , وبالرغم من ذلك قد تحتوي هذه البيانات مشاكل قد تعيق في المراحل و المهام اللاحقة. و تشمل هذه المرحلة عدة مهام أساسية وهي :

1- مهمة فهم البيانات :

ومن أهم العمليات في هذه المهمة :

- جمع البيانات : حيث تم جمع البيانات الخاصة بالمرشحين و البرامج التدريبية من المصادر المختلفة لدى الإدارة المعنية بالقرار.
- توصيف البيانات : و هي البيانات الخاصة بالمرشحين و الحقوق الخاصة بهم.
- 2- مهمة تجهيز البيانات :

ومن أهم العمليات في هذه المهمة :

- اختيار البيانات: مثل اختيار الأعمدة و الصفوف في جدول البيانات.
- تنسيق البيانات: من خلال فحص الخلايا للجدول و جعلها متناسبة مع النموذج التنبؤي المراد بناؤه , مثل أنواع البيانات (للتاريخ , المرتبة و غيرها).
- 3- مهمة تهيئة البيانات :

وذلك باختيار البيئة أو المنصة المستخدمة لإعداد البيانات , بالإضافة إلى استعراضها و معالجتها مسبقاً قبل استخدامها في المراحل اللاحقة , ومن أهم العمليات في هذه المهمة :

- شرح البيئة و البرامج المستخدمة في تجهيز البيانات و بناء النموذج لاحقاً , و هي :
- منصة google colaborytory لتنفيذ الخوارزميات الخاصة بتنقيب البيانات على البيانات عينة الدراسة.
- لغة برمجة Python لكتابة الأوامر البرمجية الخاصة بخوارزميات تنقيب البيانات.
- تحميل البيانات : و يقصد بهذه العملية تحميل ملف البيانات بعد مهمة التجهيز في منصة (google colaborytory).
- تنظيف البيانات : عزل البيانات التي تحتوي على شوائب أو قيم غير مفهومة أو مفقودة و إزالة التكرارات , و تعتبر هذه العملية مستمرة في معظم المراحل اللاحقة لضرورة موائمة البيانات لكل عملية.
- إدراج البيانات في إطار البيانات (dataframe) المراد العمل عليه.
- تكامل البيانات: و يقصد به في التجربة أنشطة دمج بعض السمات مع بعضها البعض , أو تغيير صيغة التاريخ أو استخدام بعض التعبيرات الرياضية البسيطة لاستنتاج بعض قيم المعايير و وضعها في متغير جديد لاستخدامه في المراحل و الأنشطة اللاحقة.

المرحلة الثانية : بناء النموذج التنبؤي :

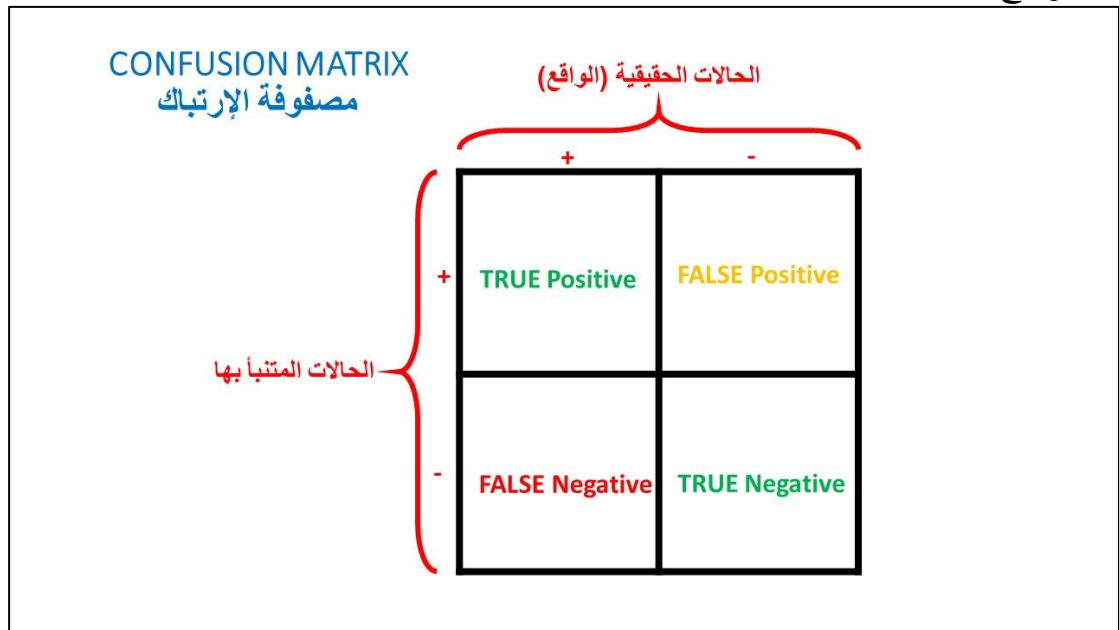
تشمل هذه المرحلة أسلوب تنقيب البيانات المستخدم و هو أسلوب التصنيف باستخدام خوارزمية شجرة القرارات و الأسلوب المناسب للتحقق من دقة النموذج التنبؤي , و تشمل هذه المرحلة عدة مهام أساسية و هي:

1- مهمة اختيار الخوارزمية :

ومن أهم العمليات في هذه المهمة :

- تعريف خوارزمية شجرة القرارات و تحديد متطلبات تطبيقها في الدراسة.
- 2- مهمة بناء النموذج التنبؤي : في هذه المهمة تم تفصيل التعليمات البرمجية لبناء النموذج و ذلك بتطبيق الخطوات التالية :

- عملية هندسة السمات (Feature Engineering) ويقصد بها موائمة المعايير الخاصة بالترشيح ليتم تطبيقها في النموذج.
- تقسيم بيانات المرشحين , إلى بيانات تدريب (training data) و بيانات إختبار (testing data) , وهي الطريقة المثلى للتعامل مع النموذج التنبؤي وتدريبه على سمات البيانات.
- تدريب النموذج التنبؤي (classifier) بواسطة البيانات التي تم تقسيمها - بيانات تدريب (training data) و بيانات إختبار (testing data) - تبعاً.
- تطبيق التعلم البرمجي الخاصة بدقة النموذج التنبؤي.
- 3- تقييم دقة النموذج التنبؤي : في هذه المهمة تم تطبيق الأساليب العلمية المتبعة لقياس دقة النموذج التنبؤي برمجياً باستخدام لغة البرمجة بايثون , وذلك بتطبيق ما يعرف بالخطأ التنبؤي وهو الفرق بين القيم الحقيقية (في الواقع) و القيم التي تم التنبؤ بها , حيث تم ذلك بناءً على عدة معايير وهي :
- الدقة (Accuracy) باستخدام مصفوفة الارتباك (Confusion Matrix) : وهي جدول يحتوي على الحالات التي تم تصنيفها بطريقة صحيحة و الحالات التي تم تصنيفها بطريقة غير صحيحة. الجدول التالي يوضح آلية عمل هذا المعيار :



شكل مصفوفة الارتباك (R. Ahmed 2019)

حيث أن :

- TRUE Positive (TP) : هي الأمثلة التي تم التنبؤ عنها على أنها موجبة وهي ضمن الصنف الحقيقي الموجب.
- TRUE Negative (TN) : هي الأمثلة التي تم التنبؤ عنها على أنها سالبة وهي ضمن الصنف الحقيقي السالب.
- FALSE Negative (FP) : هي الأمثلة التي تم التنبؤ عنها على أنها سالبة وهي ضمن الصنف الحقيقي الموجب.

FALSE Positive (FN) : هي الأمثلة التي تم التنبؤ عنها على أنها موجبة وهي ضمن الصنف الحقيقي السالب.

حيث يتم حساب الدقة (ACC) بتطبيق العملية الرياضية التالية :

$$ACC = TP + TN / (TP + TN + FP + FN)$$

- معدل الخطأ (Error Rate):

و هو تقييم النموذج عن طريق النسبة المئوية للتنبؤات غير الصحيحة بتطبيق العملية الرياضية التالية :

$$Error Rate = FP + FN / (TP + TN + FP + FN)$$

- فعالية التصنيف أو Recall أو True Positive Rate :

وهي الحالات التنبؤية الصحيحة التي صنف بصورة صحيحة على إنها موجبة و تحسب بالعملية الرياضية التالية :

$$Recall (TP Rate) = TP / (TP + FN)$$

المرحلة الثالثة : نتائج النموذج :

1- مهمة مناقشة النتائج :

في هذه المرحلة تم مناقشة النتائج المتحصل عليها من النموذج التنبؤي و تفسير البيانات و دقة النموذج بالنسبة للحالات الصحيحة والحالات الخاطئة و ذلك بحساب القيم التالية :

Precision:

و هو معدل جودة النموذج في عملية التصنيف للحالات أو معدل الحالات الصحيحة لفئة معينة, وتنبؤها المصنف بشكل صحيح , حيث دلت المتحصل عليها بأن جودة المصنف كانت بنسبة (93%) للحالات التي عددها (293 مرشح) والتي تعني عدم القبول في البرامج التدريبية والتي فئتها (0). و حيث دلت المتحصل عليها بأن جودة المصنف كانت بنسبة (87%) للحالات التي عددها (205 مرشح) والتي تعني القبول في البرامج التدريبية والتي فئتها (1). وقد تم حساب هذا المعدل بتطبيق المعادلة التالية :

$$Precision = \frac{TP}{TP+FP} = \frac{270}{270+27} = 0.93 = 93\%$$

: Recall

و هو فعالية التصنيف أو True Positive Rat وهي الحالات التنبؤية الصحيحة التي صنفبت بصورة صحيحة على إنها موجبة , حيث دلت المتحصل عليها بأن فعالية المصنف كانت بنسبة (91%) للحالات التي عددها (293 مرشح) والتي تعني عدم القبول في البرامج التدريبية والتي فئتها (0). و حيث دلت المتحصل عليها بأن فعالية المصنف كانت بنسبة (90%) للحالات التي عددها (205 مرشح) والتي تعني القبول في البرامج التدريبية والتي فئتها (1). و تحسب بالعملية الرياضية التالية :

$$Recall (TP Rate) = \frac{TP}{TP + FN} = \frac{270}{270 + 21} = 0.91 = 91\%$$

: f1-score

و هو معدل الإنسجام بين جودة النموذج و فعاليته في عملية التصنيف للحالات , حيث دلت المتحصل عليها بأن معدل الإنسجام للمصنف كانت بنسبة (92%) للحالات التي عددها (293 مرشح) والتي تعني عدم القبول في البرامج التدريبية والتي فئتها (0). و حيث دلت المتحصل عليها بأن معدل الإنسجام للمصنف كانت بنسبة (88%) للحالات التي عددها (205 مرشح) والتي تعني القبول في البرامج التدريبية والتي فئتها (1). وقد تم حساب هذا المعدل بتطبيق المعادلة التالية :

$$f1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} = 2 * \frac{0.93 * 0.91}{0.93 + 0.91} = 0.92 = 92\%$$

: accuracy

و هو معدل الدقة العامة للنموذج لجميع الحالات التي تمثلها بيانات الإختبار بما يعادل بيانات (498 مرشح) , حيث دلت المتحصل عليها بأن معدل الدقة العامة للنموذج كان بنسبة (90%). و قد تم حساب هذا المعدل بتطبيق العملية الرياضية التالية :

$$accurecy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{270 + 180}{270 + 180 + 27 + 21} = 0.90 = 90\%$$

نتائج الدراسة:

من أهم نتائج الدراسة ما يلي :

- في مهمة فهم البيانات نستنتج أن الأساس الذي تقوم عليه مشاريع اكتشاف المعرفة وتنقيب البيانات في أي منظمة يقوم على الفهم الواضح للمشكلة أو الغرض الأساسي من المشروع وذلك عن طريق جمع البيانات باستخدام عدة وسائل مثل المقابلات والاستبيانات , حيث أن فهم المجال لتحديد نوع المعرفة المطلوبة من عملية اكتشاف المعرفة في قواعد البيانات مما يسهل على المستخدمين اتخاذ القرار المناسب في اختيار وسيلة التنقيب المناسبة.
- ساهم في مهمة فهم البيانات اعتماد الأمانة على سياسات واضحة في تطبيق تخطيط وتطوير الموارد البشرية لرفع كفاءة العنصر البشري ومن أهم المظاهر لهذه السياسات وجود رؤية واضحة للأمانة فيما يتعلق بتنمية العنصر البشري والذي انعكس على الهيكل التنظيمي للإدارة العامة للموارد البشرية , من حيث وضوح المهام الرئيسية للإدارات والأقسام الفرعية بها.
- عملية جمع البيانات من الإدارة المختصة تمت عن طريق المقابلات غير المقننة مع منسوبي الإدارة , حيث تم تزويد الباحث بجميع البيانات اللازمة لتطبيق التجربة عبر البريد الإلكتروني بعد تجميعها من عدة مصادر لدى الإدارة من عدة موظفين, وهذا يفسر عدم وجود نظام مؤتمت خاص بالترشيح للبرامج التدريبية و اعتماد الإدارة على الجهود الذاتية لمنسوبيها وبالتالي من الممكن ضياع هذه البيانات بانتقال أحد المنسوين أو أي سبب آخر.
- عملية توصيف البيانات تعتبر من أهم العمليات في مهام المرحلة الأولى , حيث نتج عنها تحديد المعايير المطلوبة لتقييم المرشحين و ترميز البرامج التدريبية.
- من مهمة تجهيز البيانات تبين لدى الباحث أهمية المعرفة المسبقة للأسلوب المتبع في عملية التنقيب , حيث أن أسلوب التصنيف في التنقيب التنبؤي يتطلب المعرفة المسبقة للمدخلات والمخرجات حسب مفهوم التعلم الخاضع للإشراف (supervised learning) وهو أسلوب من أساليب التنبؤ يتلخص في تزويد النموذج التنبؤي بالمدخلات والمخرجات مسبقاً عن طريق تدريبه على هذه البيانات ثم يتم إختباره بنفس الطريقة ولكن بجزء من البيانات تم عزله مسبقاً, وهذا يتطلب الفحص والتنسيق للخلايا والصفوف والأعمدة للبيانات.
- صعوبة فهم البيانات بمعزل عن الإدارة المختصة حيث أن الأساس الذي تقوم عليه مشاريع اكتشاف المعرفة وتنقيب البيانات في أي منظمة يقوم على الفهم الواضح للبيانات أو الغرض الأساسي من المشروع.
- صعوبة جمع البيانات من الإدارة المختصة والحصول عليها من الأفراد بطريقة يدوية , والتخوف من انتهاء خصوصية الأفراد المبحوثين.
- عدم وجود نظام آلي محوسب للترشيح على البرامج التدريبية التي تطرحها أمانة العاصمة المقدسة ساهم في بقاء الحصول على البيانات , كما أدى ذلك إلى عدم الدقة في إدخال البيانات من قبل المختصين.
- نستنتج من خلال مرحلة بناء النموذج التنبؤي أن الأسلوب الأمثل للتصنيف هو أسلوب خوارزمية شجرة القرارات , حيث أن معظم المتطلبات اللازمة لتطبيقها تم توفيرها من خلال بيانات الدراسة.
- نتج عن تحديد متطلبات تطبيق خوارزمية شجرة القرارات مهام أساسية لنجاح عمل الخوارزمية مثل وضوح الفئة المستهدفة ونعني بها هنا القيم المستهدفة والتي تتمثل في قيمتين هما (Accepted = 1) و (Not Accepted = 0) وكذلك نتج عن ذلك وضوح السمات والمعايير المعتمدة لتطبيق الخوارزمية.

- نتج عن تطبيق هندسة السمات (feature engineering) وهي آلية للتعامل مع إطارات البيانات (dataframes), خط الأساس الذي سوف يتم قياس مدى دقة النموذج التنبؤي بناءً عليه.
- استعان الباحث بتقنية (one hot encoding) وهي من آليات التعامل مع إطارات البيانات (dataframes) للاستعاضة عن القيم الفئوية بقيم (0,1) لصعوبة تعامل البيئة البرمجية مع البيانات الفئوية مثل (القبول و عدم القبول و كذلك الأرقام المتسلسلة مثل 1 و 2 و 3 و ..), وهي نفس الفكرة المعتمدة في عملية ترميز متغيرات الاستبانة في برامج التحليل الإحصائي مثل برنامج (SPSS).
- تمتاز آليات عمل نماذج التصنيف بأنه يمكننا الإطمئنان مسبقاً بدقة النموذج التنبؤي من خلال تطبيق بعض المنهجيات العلمية الإحصائية والتي تم إقرارها بعد إثباتها علمياً من المختصين في الذكاء الاصطناعي و تنقيب البيانات و من أهم تلك المنهجيات منهجية (holdout method) والتي تم تطبيقها على بيانات الدراسة لضمان عدم التحيز و تحقيق العشوائية.
- بالنظر إلى نتيجة (score) والتي تمثل دقة النموذج التنبؤي نحو بيانات الاختبار (test data) والتي وصلت لنسبة (90%) وهي نسبة عالية بالنسبة للنماذج التنبؤية لأسلوب شجرة القرارات حسب النتيجة التي توصلت لها دراسة (أبو قرون 2018).
- نتج عن استخدام منهجية التحقق بالخطأ التنبؤي أو النوعي (Qualitative Error) التوصل إلى معايير مهمة للتحقق من دقة النموذج التنبؤي و مدى فعالية نتائجه, هذه النتيجة وافقت ما توصلت إليه دراسة (Kumar & Singh 2017) حيث أفادت الدراسة بأن استخدام هذه المنهجية تضمن صحة التنبؤ و من أهم المعايير التي تضمن ذلك (FN), (FP), (TN), (TP) وهي فئات الأمثلة التي تم التنبؤ بها و الناتجة من تطبيق هذه المنهجية.
- نتج عن استخدام مصفوفة الارتباك - وهي من المعايير المستخدمة في التأكد من دقة التنبؤ بمنهجية (الخطأ التنبؤي) - مخرجات ساعدت الباحث بشكل علمي و إحصائي في التأكد من سلامة و دقة التنبؤ الناتج من تطبيق النموذج بنسبة عالية استناداً إلى الدراسات العلمية و الممارسات العملية ذات الصلة بتقنيات التنقيب في البيانات و من أهم الدراسات التي وافقت ما توصل إليه الباحث دراسة (Kumar & Singh 2017), و دراسة (أبو قرون 2018), وكذلك الممارسات المعتمدة في مجتمعات التخصص في عدة منصات إجتماعية معتبرة علمياً من أهمها (Kaggle و github و stack overflow).
- نتيجة القبول و عدم القبول للمتقدمين قبل تطبيق النموذج التنبؤي و مقارنتها بنتيجة للقبول أو عدمه بعد تطبيق النموذج, حيث تظهر أن نتيجة عدم القبول تمثل (0.605618 = 60%) و نتيجة القبول (0.393482 = 40%), و مقارنتها بالنتيجة بعد التنبؤ ببيانات الاختبار حيث كانت النتيجة لعدم القبول (293 مرشح بنسبة 59%) و نتيجة القبول (205 مرشح بنسبة 41%) و الذي يوضح الدقة التي يمثلها النموذج التنبؤي المستنتج من التجربة التطبيقية حيث بلغت الدقة بنسبة تزيد عن 90%, و قابلية تطبيقه في الأنظمة التي سوف تستخدم مستقبلاً في ترشيح الموظفين على البرامج التدريبية, بالإضافة لاستخدامه بحسب البيانات المدخلة في مختلف المجالات كالتنبؤ بالأداء أو الترقيات و غيرها من أنشطة تنمية العنصر البشري في أمانة العاصمة المقدسة.
- لا يقتصر أثر تطبيق اكتشاف المعرفة من خلال تنقيب البيانات على عمليات الترشيح للبرامج التدريبية في إدارة تطوير الموارد البشرية فحسب, بل يشمل جميع جوانب و أنشطة تنمية العنصر البشري من ترقيات و قياس للأداء و عمليات التدوير و الاستقطاب, كل ما على المختصين بتنقيب البيانات فعله هو الحصول

على البيانات اللازمة وفهمها و معالجتها و إعدادها بالشكل العلمي الصحيح للإستفادة من النموذج التنبؤي المعد لهذه الأنشطة.

توصيات الدراسة:

من أهم توصيات الدراسة ما يلي :

- إنشاء مستودع للبيانات الخاصة بجميع بيانات الأمانة , الذي يفترض به أن يحتوي على عدة متاجر للبيانات لمختلف الأنشطة في الأمانة مثل المشاريع و الخدمات و الطرق و غيرها , و إنشاء قاعدة بيانات مخصصة لبيانات الموارد البشرية على وجه الخصوص.
- تصميم و تنفيذ نظام آلي محوسب للترشيح على البرامج التدريبية التي تطرحها أمانة العاصمة المقدسة لتسهيل إجراءات الترشيح على البرامج التدريبية, و ضمان دقة البيانات.
- تكثيف الجهود للتعريف بمجال تنقيب البيانات و اكتشاف المعرفة لمتخذي القرار في الأمانة و وضع برامج تدريبية للمتخصصين في تقنية المعلومات في الأمانة لتطوير قدراتهم بهذا المجال.
- تطبيق النموذج التنبؤي من خلال دمج برمجياً بأنظمة جديدة أو أنظمة قائمة مثل أنظمة الترشيح على برامج تدريبية , و أنظمة للترقيات و الإستقطاب و غيرها. و لأغراض تمكين المدراء والمسؤولين من اتخاذ قرارات أفضل تقترح الدراسة تطبيق النموذج التنبؤي في تطوير عدة أنظمة حيوية في أمانة العاصمة المقدسة.

المراجع العربية:

- احمد , أبوبكر سلطان (2017) البيانات الضخمة.. خصائصها وفرصها وقوتها, مجلة الفيصل العلمية , تاريخ الدخول 2019/11/3, متاح على : <https://www.alfaisal-scientific.com/?p=2093>
- الحسن , عمر (2019) الاستثمار في رأس المال البشري: التنمية الحقيقية, مجلة التنمية الإدارية (معهد الإدارة العامة) , تاريخ الدخول 2020/2/5, متاح على : <https://tanmia-idaria.ipa.edu.sa/Pages/%D8%A31646.aspx>
- الدوري, زكريا و أحمد , داليا (2007) دور تنقيب البيانات في زيادة أداء المنظمة دراسة تحليلية في المصرف الصناعي, المجلات الاكاديمية العلمية العراقية , تاريخ الدخول 2019/11/3, متاح على : <https://www.iasj.net/iasj?func=fulltext&ald=3348>
- الديوهجي , أبي و عبدالمحسن , عامر (2013) اكتشاف المعرفة ومشاركتها باستخدام لوحات عدادات الأعمال: دراسة حالة في كلية الإدارة والاقتصاد / جامعة الموصل, المجلة العراقية للعلوم الاحصائية, ع 13 , تاريخ الدخول 2019-10-31, متاح على : <https://www.iasj.net/iasj?func=fulltext&ald=81243>
- السلي , علي (1997) إدارة الموارد البشرية, مكتبة غريب , القاهرة , ط 1, تاريخ الدخول 2020/9/18, متاح على : https://archive.org/details/idarat_almawared_albachareya
- بن الطيب, زينب و الرباعي, سليمان (2019) الأدوار الجديدة لأخصائي المعلومات للتعامل مع البيانات الضخمة, دار جامعة حمد بن خليفة للنشر, أبو ظبي. تاريخ الدخول 2019/11/8, متاح على : https://www.qscience.com/content/journals/10.5339/jist.2018.16#abstract_content
- حسن , حسين عجلان (2008). استراتيجيات الإدارة المعرفية في منظمات الأعمال. عمان: إثراء للنشر والتوزيع.

- حنا , جبر و الكردي , فاتن (2017) تحسين اكتشاف المعرفة باستخدام تطبيقات التنقيب عن البيانات. سلسلة العلوم الهندسية تاريخ الدخول 2019/11/3 , متاح على :
<http://journal.tishreen.edu.sy/index.php/engscnc/article/view/3668>
- رابح , بلعباس و صلاح, محمد و زغبه, طلال (2017) التنقيب في البيانات واكتشاف المعرفة كآلية لتحسين جودة القرار في المؤسسة الاقتصادية, الملتقى الدولي الثاني حول التحول الرقمي للمؤسسات و النماذج التنبؤية على المعطيات الكبيرة, الجزائر, تاريخ الدخول 2019-11-10 , متاح على: <http://dspace.univ-msila.dz:8080/xmlui/handle/123456789/14013>
- الزيادات, محمد عواد (2008). اتجاهات معاصرة في إدارة المعرفة. عمان: دار صفاء سايبورال , راجيف و فرنانديز , إرما (2014). إدارة المعرفة النظم والعمليات ترجمة : مركز البحوث بمعهد الإدارة. الرياض: مركز البحوث بمعهد الإدارة.
- سيد , أحمد (2017) أدوات التنقيب عن البيانات مفتوحة المصدر دراسة تحليلية تقييمية , مجلة جامعة طيبة :للآداب والعلوم الإنسانية، العدد(7) تاريخ الدخول 2019-11-2 , متاح على:
<https://www.taibahu.edu.sa/Pages/AR/DownloadCenter.aspx?SiteId=50a13fdf-fc2d-43ef-a1d5-50bd4a749c1b&FileId=c5fec86a-2660-4160-9ef0-226e014a1fbe>
- صويص , راتب و آخرون (2011) عمليات إدارة المعرفة وتأثيرها في تحقيق الميزة التنافسية , المجلة الأردنية في إدارة الأعمال , عمان , المجلد 7, العدد 4 . تاريخ الدخول 2019/11/3 متاح على:
<https://journals.ju.edu.jo/IJBA/article/download/2722/2437>
- عليان , ريجي مصطفى (2015). إدارة المعرفة. عمان: دار صفاء.
- مقدم , عبد الحفيظ سعيد (2015). مناهج البحث في العلوم الاجتماعية و التربوية و النفسية. المجلد 1. الرياض: دار النشر الدولي.

المراجع الأجنبية:

- Avison, David & Shah, Hanifa (1997). The Information Systems Development Life Cycle , McGraw-Hill.
- Baykara, B. (2015). Impact of evaluation methods on decision tree accuracy (Master's thesis). Access date 25-10-2019. Retrieved from: <https://core.ac.uk/download/pdf/250137461.pdf>
- Gartner, Inc (2019). Big Data . access date 7-11-2019. Retrieved from: <https://blogs.gartner.com/it-glossary/big-data/>
- Gregory Piatetsky-Shapiro, and Padhraic Smyth Usama Fayyad (1996). From Data Mining to Knowledge Discovery in Databases. California : AI Magazine.
- Ikujiro Nonaka (1994). A Dynamic Theory of Organizational Knowledge Creation. Japan: Organization Science.
- Kumar, M., & Singh, A. J. (2017). Evaluation of Data Mining Techniques for Predicting Student's Performance. International Journal of Modern Education & Computer Science. Access date 20-12-2020. Retrieved from:
https://www.researchgate.net/publication/318814762_Evaluation_of_Data_Mining_Techniques_for_Predicting_Student's_Performance
- Lawal, N, Odeniyi, and Kayode, A (2015). Application of Data Mining and Knowledge Management for Business Improvement: An Exploratory Study. International Journal of Applied Information Systems

(IJASIS). Access date 8-11-2019, Available at: <https://www.ijais.org/archives/volume8/number3/715-1296>

Mehmed kantardzic (2003) .Data Mining - Concepts Models - Methods and Algorithms. John Wiley And sons, New York, USA. Access date 2-11-2019, Available at: https://www.academia.edu/18807133/Data_Mining_Concepts_Models_Methods_and_Algorithms?auto=download

Ristoski,P and Paulheim, H . Semantic Web in data mining and knowledge discovery: A comprehensive survey. Access date 12-11-2019, Available at: <https://www.sciencedirect.com/science/article/abs/pii/S1570826816000020#!>