

التنبؤ بكمية المخلفات الزراعية باستخدام تطبيق WEKA

عبد الواحد محمد أبو كريمه*^(1,2) وسامي جمعة حميدة⁽²⁾ ومحمد سعيد المرزوقي⁽²⁾ وريهام مبروك
العلمي⁽²⁾ وإبراهيم سليمان طيش⁽¹⁾

- (1). قسم الهندسة الزراعية، كلية علوم الأغذية والزراعة، جامعة الملك سعود، المملكة العربية السعودية.
(2). معهد بحوث الهندسة الزراعية، مركز البحوث الزراعية، جيزة، جمهورية مصر العربية.
(*للمراسلة: د. عبد الواحد محمد أبو كريمه. البريد الإلكتروني: aboukarima@gmail.com.)

تاريخ القبول: 2020/06/06

تاريخ الاستلام: 2020/03/29

الملخص

في هذه الدراسة تمت المقارنة بين أدوات التنقيب عن البيانات في البرنامج المفتوح المصدر WEKA للتنبؤ بكمية المخلفات الزراعية في مصر من سنة 2006 م حتى سنة 2016م، حيث تم الاعتماد على حساب كمية المخلفات بحسب الطريقة المتبعة في المنظمة العربية للتنمية الزراعية. وبينت النتائج أن أقل كمية مخلفات كان 13.57 مليون طن/سنة وأكبر كمية كانت 17.90 مليون طن/سنة. تم ترتيب أدوات التنقيب عن البيانات المستخدمة بحسب أقل قيمة لمتوسط انحراف الخطأ، وتبين أن الأداة المسماة جدول القرار (Decision Table) هي أفضل أداة للتنبؤ بكمية المخلفات الزراعية، حيث قيمة متوسط انحراف الخطأ لها كانت -9.149 ألف طن/سنة، كما تبين أن قيم معامل التحديد (R^2) لجميع الأدوات فوق 95%، وكان المتوسط الحسابي لكمية المخلفات المتنبأ بها من 17 أداة للتنقيب عن البيانات متقارب، حيث أقل متوسط حسابي كان 2502.17 ألف طن/سنة وأكبر متوسط حسابي كان 3579.66 ألف طن، بينما المتوسط الحسابي للقيم الفعلية في بيانات الاختبار كان 3300.06 ألف طن/سنة. واستنادًا لذلك، وضعت التوصية الخاصة بالبحث وهي ضرورة الاهتمام الكافي بالتنبؤ بالمخلفات الزراعية من خلال استخدام النماذج الرياضية، حيث يمكن إدارة هذه المخلفات والاستفادة منها في إنتاج الطاقة أو السماد أو كعلف حيواني.

الكلمات المفتاحية: تعلم الآلة، المخلفات الزراعية، النمذجة، البيئة.

المقدمة:

يترتب على تراكم الكميات الضخمة من المخلفات الزراعية، مع عدم الاستفادة منها، انتشار الميكروبات والأمراض المعدية، مما يؤدي في النهاية إلى حدوث خلل في النظام البيئي داخل المزرعة ككل، وينعكس ذلك على الأفراد، بالإضافة إلى أن التعامل غير السليم مع المخلفات الزراعية كحرقها، يؤدي إلى تلوث البيئة وحدوث مشكلات صحية للإنسان، وأيضًا تخزين هذه المخلفات المزرعية بالطرق التقليدية، يؤدي إلى انتشار الحشرات والفئران والكائنات الحية الدقيقة غير المرغوب فيها والمسببة للعديد من الأمراض، بالإضافة إلى

نشوب الحرائق (مكاوي، 2010م). وفي المقابل يمكن التعامل مع هذه المخلفات بما يحقق الاستفادة منها من خلال تحويل المخلف من صورة إلى أخرى أو بيعه أو الاستفادة منه كما هو (إمام، 2007 م). ويضيف الحجار (2003 م) أنه يجب الاستفادة من المخلفات الزراعية من خلال منظومة متكاملة تتضمن تحويل هذه المخلفات إلى سماد أو وقود أو علف حيواني. كما أن المخلفات الزراعية تمد العالم بما مقداره 33% من الوقود الحيوي (Yevich and Logan, 2003).

يتمثل السعي الدائم في رفع إنتاجية المحصول لتوفير المزيد من الغذاء ليقابل التزايد المستمر في أعداد السكان وارتفاع مستوى المعيشة، إلى مزيد من الاستنزاف للموارد المتاحة، والمغالاة في استخدام التقنيات الحديثة (الميكنة والأسمدة الكيماوية والمبيدات، ومنظمات النمو... الخ)، والاستخدام غير الرشيد لمياه الري وما يصاحب ذلك من ممارسات خاطئة تسبب الكثير من الأضرار بالبيئة الزراعية. وينقسم الإنتاج الزراعي لشقين أساسيين متلازمين، وهما الإنتاج النباتي والإنتاج الحيواني، وعلى ذلك تتكون المخلفات النباتية والحيوانية، حيث تتراوح كمية هذه المخلفات سنوياً بمقدار من 30 إلى 35 مليون طن (Abou Hussein and Sawan, 2010)، والمخلفات الحيوانية وحدها تمثل حوالي 14.6 مليون طن/سنة (Abou El- Azayem and Abd El-Ghani, 2010)، كما بينا الشايب ومحمد (2011م) من خلال جهاز شؤون البيئة في مصر أن المخلفات الزراعية تمثل حوالي 32% من إجمالي المخلفات بمصر لسنة 2011م.

توجد معوقات وتحديات تقف أمام استغلال المخلفات الزراعية والاستفادة، منها ما هو معوقات فنية مثل طبيعة ونوعية المخلفات وكيفية معاملتها (المنظمة العربية للتنمية الزراعية، 2006م)، ومنها ما يتعلق بكيفية اتخاذ القرار بشأنها، ومن توصيات المنظمة العربية للتنمية الزراعية بالنسبة للمخلفات الزراعية هو إعداد قاعدة بيانات ومعلومات خاصة بالكمية ونوعية المخلفات. وترى سردينة (2004 م) أنه بالرغم من وجود عديد من التقنيات الحديثة التي يمكن تطبيقها لتدوير هذه المخلفات، وتعظيم الاستفادة منها، إلا أنه لا يمكن الاعتماد على وجود التقنيات وحدها، لأن ذلك لن يقود إلى التقدم، ومن أجل ذلك فلا بد للمؤسسات الإعلامية والثقافية والتعليمية والصحية أن تقوم بنشر الوعي حول عناصر البيئة وأهمية المحافظة عليها، وكذلك حسن التعامل والاستفادة من المخلفات، لأن مشكلة تلوث البيئة، والنتيجة عن التخلص غير الآمن للمخلفات هي في الأساس مشكلة سلوك إنساني. وفي دراسة عن أنواع المخلفات الزراعية والسمات البيئية والكيميائية لها والعلاقة بينها وبين تقنية العمارة الذكية، وجد أنه يمكن الاستفادة منها بتقنية إعادة التدوير للحفاظ على الطاقة وترشيد استخدامها داخل المباني (نصير، 2018م). وهناك تقديرات للمخلفات الزراعية، ومعظم الدراسات في هذا الجانب تكون في الإحصاءات الزراعية الحكومية، والتي تعطي معلومات عن كمية الإنتاج والمساحة لهذه المحاصيل وهناك علاقة بين كمية الإنتاج ومقدار المخلفات الناتجة (Yevich and Logan, 2003).

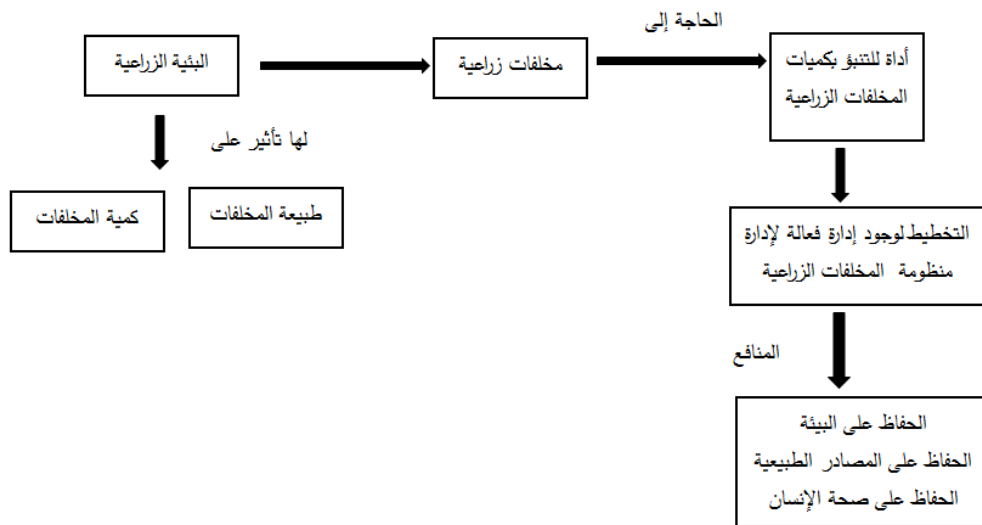
تعتبر عملية التقيب عن البيانات هي عملية تحليل البيانات من منظور مختلف واستخلاص علاقات بينها وتلخيصها إلى معلومات ذات فائدة من خلال استعمال مجموعة من الأدوات المعقدة، وبعض من هذه الأدوات تشمل أدوات الإحصاء الاعتيادية والذكاء الاصطناعي (علي، 2018م). ويمكن تقسيم أدوات التقيب عن البيانات إلى أدوات التقيب عن البيانات المباشرة وغير المباشرة؛ حيث يتمثل الهدف من التقيب عن البيانات المباشرة في استخدام البيانات المتاحة لإنشاء نموذج مع وصف للمتغيرات؛ أما الهدف من التقيب عن البيانات غير المباشرة هو عدم توافر اختيار لمتغير محدد، لكي يتم بناء علاقة بين كل المتغيرات. وتدرج أدوات التصنيف

والتقدير والتنبؤ ضمن أدوات التنقيب عن البيانات المباشرة؛ أما أدوات الارتباط والعنقدة والوصف والعرض فتندرج ضمن التنقيب عن البيانات غير المباشرة (سيد، 2015م).

تتمثل أهمية الدراسة من الناحية النظرية محاولة للإسهام في الجهود العلمية الهادفة إلى استخدام أدوات التنقيب عن البيانات في المساعدة في اتخاذ القرار بالنسبة للمخلفات الزراعية، كما أن هذه الدراسة بما اتبع فيها من منهج علمي وما تضمنتها من مفاهيم وما استخدم فيها من أساليب تحليل البيانات، وما قد يسفر عنه من نتائج وتوصيات يمكن أن يساهم في فتح آفاق جديدة لإجراء المزيد من البحوث التي تستهدف المخلفات الزراعية، وذلك للحفاظ على البيئة، لذا فإن البحث عن استراتيجيات لحماية البيئة والحفاظ عليها يصبح في حد ذاته عملاً له أهميته البالغة في مواجهة أو التخفيف من حدة المشكلات البيئية (سردينه، 2004؛ حبيب ورشاد، 2016م). ومن الناحية التطبيقية، فإن هذه الدراسة، وما يمكن أن تسفر عنها من عرض لأدوات التنقيب عن البيانات التي تتعلق بطرق التنبؤ والتصنيف للمخلفات الزراعية تعتبر أساس يمكن تطبيقه بطرق أخرى للحفاظ على البيئة، حيث يمكن استخدام هذه الأدوات من خلال بناء نماذج رياضية تساعد المنظمات على اتخاذ قرارات (الدوري وأحمد، 2007م).

في ظل التطور المتسارع لتكنولوجيا الحاسوب والمعلوماتية، فإن كمية البيانات التي يمكن تخزينها يتعاظم بشكل كبير، ولا ينتهي المطاف بالقدرة على تخزين هذه البيانات، حيث تأتي الخطوة الأهم، وهي كيفية استثمار هذه الكميات الضخمة من البيانات (الجناعي وآخرون، 2011م). ويمكن القول إن هذه البيانات تمثل ثروة حقيقية يمكن أن تهمل، كما يمكن أن تستثمر في تحليل ما مضى من أحداث، ومن ثم يمكن أن يتم رسم السياسات والخطط بشكل واضح، يتيح التنبؤ بفرص التحسين المستقبلي للخطط الموضوعة من قبل منظمة ما. ومع هذه الكميات الضخمة من البيانات، فإن الطرائق التقليدية لتحليل البيانات، والتي هي مزيج من الطرائق الإحصائية وبعض النظم الحاسوبية المصممة لإدارة قواعد البيانات، باتت تعاني الكثير من المشكلات في التعامل مع هذا النوع من البيانات، بالإضافة إلى ذلك، فإن تلك الطرائق التقليدية تعتمد بشكل كلي على القدرات الذهنية والفنية للقائمين عليها، ومن ثم على خبرة محلل البيانات في توجيه التحليل لاستخراج مؤشرات مفيدة وعميقة لصناع القرار؛ حيث يفترض المحلل فرضيات سابقة للعلاقات الموجودة بين المتغيرات المختلفة للبيانات وجل ما يفعله بعد ذلك هو استخدام التحليل لبرهنة أو دحض هذه الفرضيات (الجناعي وآخرون، 2011م).

تتمثل المشكلة الرئيسية التي تواجه صانعي القرار في إدارة المخلفات الزراعية في كيفية التنبؤ بكمياتها المحتمل تواجدها في المستقبل القريب، وذلك بهدف وضع استراتيجية أكثر ملائمة لإدارتها، ويمكن التوصل إلى ذلك من خلال النمذجة التنبؤية (شاهين وآخرون، 2017م). وأن التنبؤ بالمخلفات الصلبة وإيجاد إطار مستقبلي للتنبؤ يكون أداة فعالة للتخطيط المستقبلي بكمياتها (Abbas and El Maddison, 2016)، كما أن التنبؤ بالمخلفات يفيد في حساب الكميات اللازمة لإنتاج الطاقة الحيوية منها (Hanandeh, 2016). ومن جهة أخرى مازال صانع القرار في مجال إدارة المخلفات يحتاج إلى مؤشرات وبيانات ضرورية للتحليل الاقتصادي، لاختيار الطريقة الحكومية المناسبة للتعامل مع المخلفات (Loukil and Rouached, 2020). وإن الحاجة لوجود منظومة لإدارة المخلفات المزرعية مهمة بسبب تأثيرها في البيئة، وهذه المنظومة يمكن أن تكون خطواتها، كما في الشكل (1)، كما ورد في دراسة (Heilala, 2018)، حيث قرر الباحث أن واحد من أهم هذه الخطوات هي عملية التنبؤ.



الشكل 1. منظومة مقترحة لإدارة المخلفات الزراعية، معدلة من قبل (2018) Heilala.

ولأهمية التنبؤ، قام Achinas *et al.*, (2016) بإجراء تحليل نظري لإمكانية التنبؤ بكميات الغاز الحيوي من المخلفات الزراعية وفي دراسة Jiang *et al.*, (2020) تم استخدام أدوات التقيب عن البيانات في تصميم إطار تحليلي لتحليل سلوك كميات المخلفات المنزلية في شنغهاي بالصين، والهدف هو إيجاد وسيلة لمتخذي القرار لإدارة عمليات تجميع المخلفات المنزلية الصلبة لوضع الخطط المستقبلية. بينما بين كلا من Kaewdiew *et al.*, (2019) أنه من المهم حصر كميات المخلفات الزراعية، وتحديد النسبة من المحاصيل التي تنتج مخلفات زراعية، حتى يمكن إنتاج الغاز الحيوي، وبينوا أنه لمحصول الذرة فإن نسبة المستفاد من المحصول هو فقط 41.49% والباقي مخلفات، ولقصب السكر نسبة المستفاد منه 66.53% والباقي مخلفات ومحصول الأرز، نسبة 70.16% هي التي يستفاد منها والباقي مخلفات، كما أجرى الباحثون أنفسهم طريقة للتنبؤ بكميات المخلفات الزراعية الغير مستخدمة في إنتاج الغاز الحيوي في تايلاند للفترة 2015-2022م، كما أن كمية المخلفات الزراعية الناتجة من محصول ما، تعتمد على الظروف المناخية والأسمدة المستخدمة والصنف المنزرع بحسب ما ذكره Arnard and Kristoferson (1985)، وهناك نسب ما بين الكمية المنتجة من المحصول ونسبة المخلفات الناتجة منه، كما هو موضح بالجدول (1).

الجدول 1. نسبة المخلفات إلى المحصول.

المحصول	القيمة المختارة	المدى (Yevich and Logan, 2003)	Barnard and Kristoferson (1985)
القمح	1.3	0.9-1.6	0.7-1.8
الذرة	2.0	0.9-4.0	1.2-2.5
قش الأرز	1.5	0.8-2.5	1.1-2.9
الشعير	1.6	1.4-2.0	0.9-1.8
قصب السكر	0.15	0.05-0.2	-
القطن	4.0	3.0-5.5	3.5-5.0
الفول السوداني	0.4	0.25-0.5	0.5

بين (Kumar and Samadder 2017) أن التنبؤ الدقيق لكمية المخلفات الصلبة المنزلية مهم جدًا للإدارة الفعالة لها، والنماذج الرياضية الناتجة غالبًا ما تكون مفيدة للتنبؤ الدقيق بمعدل توليد المخلفات الصلبة المنزلية، هذه النماذج تعتمد على مدخلات مثل حجم المنزل ودخل الأسرة ومستوى التعليم والوقود المستخدم في المطبخ، حيث تم استخدام الانحدار الخطي المتعدد في توليد مثل هذه النماذج. وفي نفس السياق كان هدف دراسة (Kannangara et al., 2018) هو تطوير نماذج تنبؤيه لكمية المخلفات الصلبة المنزلية باستخدام أدوات تعلم الآلة مثل شجيرات القرار (decision trees)، والشبكات العصبية (neural networks)، وبينت النتائج أنه يمكن لأدوات تعلم الآلة التنبؤ بهذه الكميات من خلال متغيرات ديموجرافية demographic ومتغيرات مرتبطة بالإنسان Socio-economic في أونتاريو، كندا بنسبة 72%. وفي دراسة (AI-Subu 2015) لتطوير نموذج رياضي لتوقع مكونات النفايات الصلبة ومعدل توليدها، وتخوفات السكان تجاه مرافق إدارة النفايات الصلبة في محافظتي نابلس وجنين، فلسطين، تم الاستعانة باستبيان لتجميع البيانات اللازمة، حيث تم اشتقاق سبعة نماذج رياضية من خلال تحليل الانحدار المتعدد لتقدير كمية النفايات الصلبة المتولدة يوميًا، وأظهرت المؤشرات الإحصائية أن النماذج كانت عالية التنبؤ، ومن فوائد النموذج الرياضي أنه قد يساعد اصحاب القرار لوضع خطط أفضل لإدارة النفايات ومرافقها. وبينت الباحثة أن من خلال استخدام النموذج الرياضي، يمكن المساهمة في تقليل النفايات، وتحسين كفاءة استخدام الموارد من خلال معرفة طبيعة النفايات التي يتم إنتاجها بواسطة الأفراد، حيث يؤثر نوع النفايات على الصحة العامة والبيئة المحيطة، وبالتالي فإن الحصول على معلومات جيدة متعلقة بمكونات النفايات يساعد أيضًا في تحسين التخطيط وتحسين إدارة مرافق النفايات الصلبة، مثل تحسين فعالية منشآت إعادة التدوير، عندما يتم معرفة مكونات النفايات التي يمكن إعادة تدويرها. وفي دراسة (Heilala 2018) بين أن الكميات الكبيرة من المخلفات البشرية تتطلب إدارة فعالة للتعامل بكفاءة مع هذه الكميات المتزايدة، ويتأتى ذلك من خلال استخدام أساليب تحليل البيانات أثناء جمع هذه المخلفات، حيث استخدم في الدراسة تقنيات التنقيب عن البيانات واكتشاف المعرفة في تحليل بيانات وزن المخلفات المنزلية المختلطة، وتمت معالجة البيانات مع خطوات اكتشاف المعارف التقليدية، بما في ذلك اختيار تصفية البيانات، ومعالجة البيانات المسبقة، وتحليل البيانات، والتحقق من صحة وتفسير النتائج وبينت النتائج أن تقنيات التنقيب عن البيانات واكتشاف المعرفة قادرة على التنبؤ باتجاهات الزيادة في المخلفات البشرية في منطقة هلسنكي، فنلندا. وفي دراسة (Abbasi and El Hanandeh 2016) اختبرا أربعة طرق من أدوات الذكاء الاصطناعي للتنبؤ بالمعدل الشهري للمخلفات الصلبة في استراليا، هذه الطرق شملت آلة المتجهات الداعمة (Support Vector Machine)،

ومنظومة استنتاج عصبية ضبابية مكيفة (ANFIS)، الشبكات العصبية الاصطناعية، وطريقة الجار الأقرب، وبينت النتائج أن كل الطرق صالحة للتعامل مع البيانات، ولكن طريقة ANFIS كانت هي الأدق في التنبؤ. ومن خلال استعراض الدراسات وأهمية المخلفات الزراعية في إنتاج الطاقة أو السماد أو كأعلاف حيوانية، لذا تهدف الدراسة إلى التعرف على أدوات التنقيب عن البيانات التي يمكن الاستفادة منها في تحليل البيانات، وبالتحديد في التنبؤ بكمية المخلفات الزراعية والذي يساهم في إدارة المخلفات الزراعية.

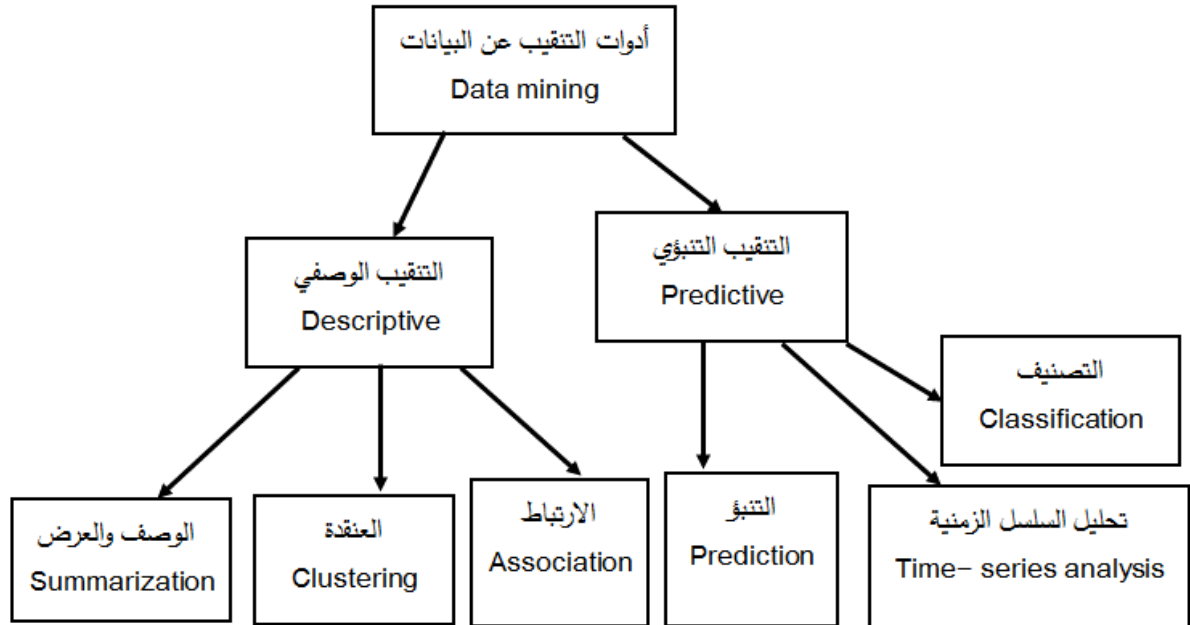
مواد البحث وطرقه:

نبذة عن التنقيب في البيانات:

تعتبر عملية التنقيب عن البيانات هي عملية البحث في قواعد البيانات (مستودعات البيانات) عن معرفة غير مكتشفة وغير متوقعة أي الحصول على معرفة جديدة غير موجودة في قواعد البيانات الأصلية وتكون هذه المعرفة مهمة بحيث تساعد في اتخاذ القرار، ويعد التنقيب في البيانات مرحلة من مراحل استكشاف المعرفة في قواعد البيانات والتي تشير إلى استكشاف الأنماط (الجناعي وآخرون، 2011م). من خلال بعض المراحل، حيث تتمثل المرحلة الأولى في اختيار البيانات وهي مرحلة اختيار البيانات المرشحة للدراسة من مستودع البيانات الكلي بحسب الغرض من الدراسة، ثم المرحلة الثانية وهي مرحلة تنقية البيانات وهي مرحلة تنقية البيانات التي تحتوي على تشويش أو قيم شاذة من مجموعة البيانات بحيث يتم الحصول على قاعدة بيانات نقية، ثم المرحلة الثالثة وهي مرحلة تحويل البيانات، وفيها يتم تحويل البيانات التي تم اختيارها إلى شكل ملائم لخوارزميات والتطبيقات التي ستستخدم في الدراسة. حيث قد تشترط بعض الخوارزميات وجود البيانات على هيئة معينة قبل تطبيقها ويلي ذلك المرحلة الرابعة وهي مرحلة التنقيب في البيانات، وفيها يتم تطبيق طرق وخوارزميات ذكية وبسلسل مناسب لاستكشاف أنماط مفيدة، ثم المرحلة الأخيرة وهي مرحلة التقييم، فبعد استكشاف الأنماط المهمة والتي تمثل المعرفة يتم تقييمها بطرائق كمية ووصفية مختلفة وملائمة للتقنيات التي طبقت وباستخدام مقاييس محددة في بيئة المشكلة.

تهتم عملية التنقيب عن البيانات بعمليتين أساسيتين (الشكل 2) متمثلتين في التنقيب الوصفي والتنقيب التنبؤي، حيث في التنقيب الوصفي يعتمد تحليل البيانات على إعادة تنظيم البيانات والتنقيب فيها لاستخراج المعلومات منها، أما التنقيب التنبؤي يعتمد تحليل البيانات على استخدام معلومات سابقة لتوقع ما سيحدث في المستقبل، ويعد تحليل السلاسل الزمنية والتصنيف والانحدار أنواع أساسية للتنقيب التنبؤي، فيستخدم التصنيف للتنبؤ بالقيمة المنفصلة أو الرمزية (المتغير الوصفي)، ويستخدم التصنيف بشكل واسع في تحليل مجموعة من البيانات ووضعها على شكل أصناف أو أقسام يمكن استخدامها فيما بعد لتصنيف البيانات مستقبلاً، فالتصنيف يقصد به تقسيم البيانات إلى مجاميع، أما الانحدار فيستخدم للتنبؤ بالقيم المستمرة (المتغير الكمي)، حيث يعتبر المتغير مقدار له خصائص رقمية (كمية) وغير رقمية (وصفية). ويمكن تقسيم أدوات التنقيب عن البيانات إلى أدوات التنقيب عن البيانات المباشرة وغير المباشرة؛ حيث يتمثل الهدف من التنقيب عن البيانات المباشر في استخدام البيانات المتاحة لإنشاء نموذج مع وصف للمتغيرات؛ أما الهدف من التنقيب عن البيانات غير المباشر هو عدم توافر اختيار لمتغير محدد، لكي يتم بناء علاقة بين كل المتغيرات. وتدرج أدوات تنقيب البيانات مثل التصنيف والتنبؤ وتحليل السلاسل الزمنية ضمن أدوات التنقيب عن البيانات المباشر؛ أما أدوات التنقيب الوصفي مثل

الارتباط والعنقدة والوصف والعرض فتندرج ضمن أدوات التنقيب عن البيانات غير المباشر (سيد، 2015م).



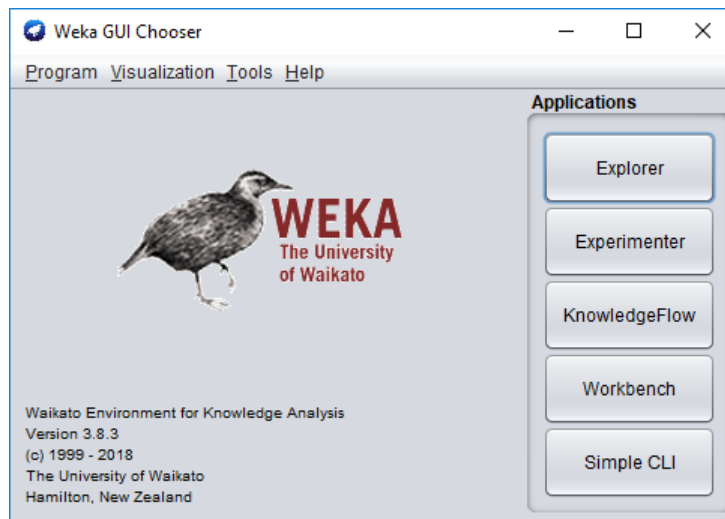
الشكل 2. مهام التنقيب عن البيانات، منقحة من سيد (2015م).

يعتبر أسلوب التحليل العنقودي من الطرق الاحصائية المهمة والمستخدم في التنقيب الوصفي، والتي يعتمد على تحليل متغيرات محددة تعتمد على نقاط التشابه والاختلاف بين البيانات. والتحليل العنقودي يستخدم لتصنيف ودراسة تجمعات البيانات أو المشاهدات أو العناصر في مجاميع متجانسة فيما بينها ومختلفة عن باقي المجاميع بالاعتماد على مجموعة من الصفات أو المتغيرات وإن عملية التعتد تعني وضع العناصر في مجاميع، أي انها تمثل الحصول على مجموعة واحدة أو أكثر من المجاميع التي تضم عدداً من العناصر على درجة كبيرة من التشابه فيما بينها، وعملية الحصول على هذه المجاميع يتم عن طريق تجزئة العدد الكلي للعناصر الى مجاميع او كل مجموعة تضم عدد من العناصر بالاعتماد على مصفوفة التشابه. ومن أدوات المستخدمة في التنقيب الوصفي قواعد الارتباط، وهي إحدى التقنيات الواعدة في أدوات التنقيب عن البيانات كأداة من أدوات استكشاف المعرفة، ولديها القدرة على معالجة أعداد هائلة من البيانات، وتسمح باستنتاج كل القوانين الممكنة التي تشرح بعض الصفات الموجودة اعتماداً على الأخرى (An et al., 2003)، ويمكن تطبيق عدة أدوات للتنقيب عن البيانات في الزراعة من خلال تطبيقات مختلفة Kodeeshwari and Ilakkiya, (2017).

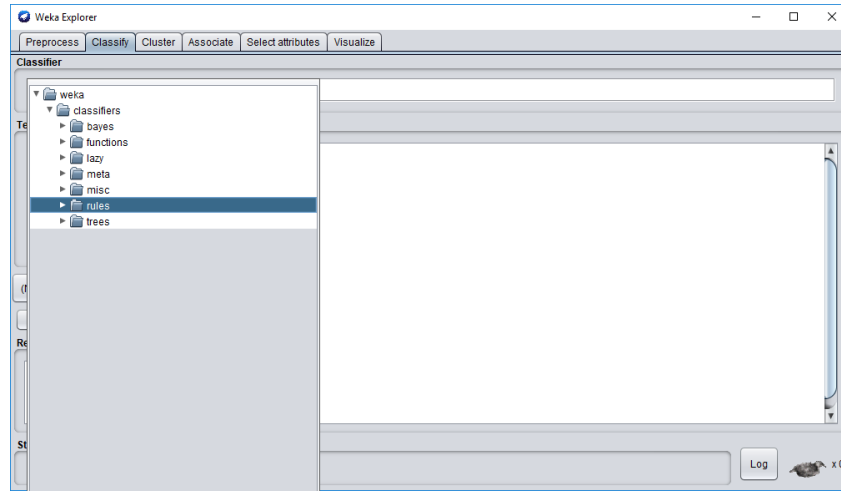
البرنامج المستخدم:

التنقيب عن البيانات هو مجال مهم جداً، ينطوي على تحليل مجموعات كبيرة من البيانات، من أجل الكشف عن أنماط، واساليب جديدة بهدف معالجتها بشكل احترافي من خلال خوارزميات معقدة يوفرها التطبيق الحاسوبي WEKA، وهذا التطبيق هو مجموعة من الأدوات والبرمجيات المتقدمة التي تساعد المستخدم في التنقيب عن البيانات وتجميعها وتحليلها باحترافية من خلال خوارزميات معقدة كتعلم الآلة وغيرها وكل هذا يساعد على معالجة البيانات بشكل جيد (Hall et al., 2009). تشمل حزمة WEKA خمسة تطبيقات رئيسية، منها المستكشف Explorer لاستكشاف البيانات، والمجرب للقيام بالتجارب والاختبارات الإحصائية 'Experimenter' وتدقق المعرفة Flow Knowledge، وواجهة الأوامر لتنفيذ أوامر Simple CLI Weka، و Work bench، (الشكل 3). وتسمح

أدوات WEKA بفتح مجموعة من البيانات وتحريها، كما يمكن تغيير محتويات البيانات، وتغيير السمات، وتصنيف البيانات المتاحة، وفقاً لمجموعة محددة مسبقاً من القواعد، علاوة على إمكانية إجراء عملية التقييم لأدوات تعلم الآلة Machine Learning المستخدمة وتشرط هذه الحزمة تحميل ملف البيانات بطريقة معينة، حيث تعلم الآلة هو علم من علوم الذكاء الاصطناعي وهو قدرة الآلة على التعلم، وإنتاج معلومات جديدة من خلال الملاحظة والاستنتاج، ومقارنة الخبرات والمعلومات المخزنة. وهناك عدد من الطرق التي يمكن استخدامها في هذه الحزمة باستخدام خوارزميات مختلفة في البنيات الرياضية، ويعد WEKA من أشهر أدوات برمجيات التنقيب عن البيانات مفتوحة المصدر، والتي تتكون من العديد من أدوات التنقيب عن البيانات (سيد، 2015م) والتي توفر واجهة بسيطة سهلة التعلم للوصول لأدوات التنقيب عن البيانات. والشاشة الرئيسة هي شاشة Explorer (الشكل 4)، وهي الشاشة التي يتم من خلالها إدخال البيانات واختيار أداة التنقيب المناسبة وأدوات التنقيب للتنبؤ في WEKA تشمل ستة أقسام هي meta, lazy, functions, misc, rules و trees، وكل قسم يندرج تحته عدة أدوات، ولمعرفة البنية الرياضية لكل أداة يمكن الرجوع إلى المصادر الموجودة تحت كل أداة معروضة في التطبيق الحاسوبي WEKA.



الشكل 3. واجهة التطبيق الحاسوبي WEKA.



الشكل 4. الشاشة الرئيسية في التطبيق الحاسوبي WEKA.

منهجية الدراسة والبيانات المستخدمة:

تم اعتماد منهجية تتكون من عدة مراحل تتلخص في جمع البيانات (من الكتاب السنوي للإحصاءات الزراعية العربية، المنظمة العربية للتنمية الزراعية، جامعة الدول العربية، مجلدات مختلفة، أرقام 33،34،35،36،37)، حيث تم الحصول على بيانات توضح إجمالي الحبوب، حيث شملت الحبوب محاصيل القمح، الشعير، الذرة الشامية، الذرة الرفيعة والأرز، والدرنات والجذور، قصب السكر، بنجر السكر، عباد الشمس والسمسم كإحصاءات زراعية في جمهورية مصر العربية من سنة 2006 م حتى سنة 2016 م لاستخدامها في الدراسة، حيث رتبنا هذه البيانات في جدول أكسل. وتم حساب كمية المخلفات لهذه الأنواع من المعادلة (1) والمعادلات القياسية المقابلة، والتي وقدرت هذه الكميات على أساس المادة الخام المنتجة وليس على أساس المادة الجافة، كما تم الاعتماد على المعدلات القياسية لتقدير المخلفات على أساس نسبة الغذاء الصالح للإنسان، للدرنات 1:1.5، للحبوب 1:2، للحبوب الزيتية 1:6 وللمحاصيل السكرية 1:10، كما تمت الحسابات بالطريقة التالية (الكتاب السنوي للإحصاءات الزراعية العربية، 2017):

$$\text{كمية المخلفات (كجم/سنة)} = \text{كمية الإنتاج (كجم/هـ)} \times \text{المساحة (هـ/سنة)} \times \text{المعدل القياسي} \dots (1)$$

تم الاعتماد على هذه الطريقة لإيجاد كمية المخلفات الزراعية في جمهورية مصر العربية من سنة 2006 م حتى سنة 2016 م واستخدمت الطرق المختلفة لأدوات التنقيب عن البيانات في التنبؤ بهذه الكميات، وتم استخدام برنامج WEKA في تنفيذ تحليل البيانات، حيث يوفر هذا البرنامج منصة تفاعلية لإدخال البيانات وإظهار النتائج وتقييم أداء كل أداة من خلال معايير إحصائية. وتمت المعالجة الأولية للبيانات على الشكل الذي يناسب التطبيق الحاسوبي المستخدم، ثم مرحلة تطبيق تقنيات التنقيب عن البيانات على البيانات، ثم مرحلة التقييم لاختيار أفضل أداة للتنبؤ بكمية المخلفات الزراعية. في هذه الدراسة تم تقسيم البيانات إلى 83% منها للتدريب و17% منها للاختبار، حيث لم يتم تغيير متغيرات أدوات التنقيب عن البيانات في التطبيق وتم تركها كما هي Default. كانت المدخلات 7 مدخلات ومخرج واحد. تم إعطاء نوع المحصول رقم ما بين صفر وواحد، حيث إذا كان واحد فيكون باقي المحاصيل صفر وهكذا. وتم تعديل قيم المساحة والإنتاجية وكمية الإنتاج وكمية المخلفات لتكون بين قيم 0.15 و0.85، حتى يمكن رسم العلاقة بين كل من كمية المخلفات والمساحة، وكمية المخلفات والإنتاجية لمعرفة اتجاه كمية المخلفات بالنسبة للمساحة وكمية المخلفات، باستخدام المعادلة التالية:

$$X = \frac{(t - t_{\min})}{(t_{\max} - t_{\min})} \times (0.85 - 0.15) + 0.15 \dots\dots\dots (2)$$

حيث X هي القيمة المعدلة، و t هي القيمة الأصلية، و t_{\min} هي أقل قيمة، و t_{\max} هي أكبر قيمة (Vesta Services, 2000).

تقييم أدوات التنقيب عن البيانات

في هذه الدراسة، تم تقييم أدوات التنقيب عن البيانات المستخدمة، ومن ثم ترتيبهم على حسب القدرة التنبؤية لكل منها، من خلال عدة معايير إحصائية منها متوسط انحراف الخطأ (Mean Bias Error (MBE)، ومتوسط الجذر التربيعي لمربع الخطأ (Root Mean Square Error (RMSE)، كما يلي (Ogolo, 2010):

$$MBE = \frac{1}{n} \sum_{i=1}^n (FC_{est} - FC_{obs}) \dots\dots\dots (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (FC_{est} - FC_{obs})^2} \dots\dots\dots (4)$$

حيث FC_{est} و FC_{obs} هما كمية المخلفات الفعلية والمنتبأ بها على الترتيب، و n عدد القراءات في بيانات مجموعة الاختبار، حيث أن النموذج الذي يعطي قيمة أصغر لمتوسط انحراف الخطأ ومتوسط الجذر التربيعي لمربع الخطأ، يكون النموذج هو أفضل نموذج، كما أن القيمة الموجبة لمتوسط انحراف الخطأ تعني أن النموذج يتنبأ بقيمة أكبر من الفعلي، بينما القيمة السالبة لمتوسط انحراف الخطأ تعني أن النموذج يتنبأ بقيمة أقل من الفعلي (Kolebaje et al., 2012). تم حساب معامل التحديد R^2 لتوضيح العلاقة الخطية بين القيم الفعلية والقيم المنتبأ بها. وتم استخدام درجة الرتبة لترتيب الأدوات بحسب المعادلة التالية:

$$Rank \quad Score = \frac{Abs(MBE)}{Mean} + \frac{RMSE}{Mean} + T \dots\dots\dots (5)$$

حيث Mean هو متوسط القراءات المنتبأ بها و T قيمة إحصائية، يمكن حسابها كما يلي (Khan et al., 2012):

$$T = \sqrt{\frac{(n-1)MBE^2}{(RMSE^2 - MBE^2)}} \dots\dots\dots (6)$$

والنموذج الذي له رتبة أقل هو أفضل نموذج (أداة التنقيب) بحسب (Kolebaje et al., 2016).

النتائج والمناقشة:

تحليل البيانات:

استخدم برنامج أكسل لحساب عدة معايير إحصائية لتحليل البيانات، وهي أقل قيمة وأكبر قيمة والانحراف المعياري والمتوسط الحسابي، وقيم الالتواء والتلطح، ومن الانحراف المعياري والمتوسط الحسابي تم حساب معامل الاختلاف، وكلما كانت قيمة معامل الاختلاف منخفضة دل ذلك على تكتل البيانات حول متوسطها وعدم تشتتها والعكس صحيح، ويعتبر التوزيع التكراري المتماثل هو ما كانت تكراراته موزعة توزيعاً متماثلاً حول وسطه الحسابي بما يعني قيم المتغير المتساوية البعد عن الوسط الحسابي لها نفس التكرار (محمد، 2007م). تم حساب الالتواء Skewness، وهو مقياس نسبي يمكن استخدامه للمقارنة بين التواء التوزيعات، كما أن قيمة مقياس الالتواء محصورة بين +3، و -3 فالقيمة السالبة تعني الالتواء جهة اليسار والقيمة الموجبة تعني الالتواء جهة اليمين والقيمة صفر تعني عدم وجود التواء أو التماثل (التوزيع متماثل) (محمد، 2007م). أما التلطح Kurtosis فهو مقدار التدبب (الارتفاع أو

الانخفاض) في قمة المنحنى مقارنة بقيمة منحنى التوزيع الطبيعي. كما أن التقلطح Kurtosis، وهو مقياس لتكرارات القيم على طرفي توزيع المتغير، فإذا كانت قيمة التقلطح كبيرة كانت تكرارات القيم أكبر على طرفي التوزيع، أما إذا كانت تكرارات القيم أقل على طرفي التوزيع كان التقلطح أقل. وتكون قيمة معامل التقلطح صفر في حالة التوزيع الطبيعي المعياري. ففي حالة ما يكون معامل التقلطح للبيانات الأصلية أكبر من 3 يكون المنحنى مدبب لأعلى بمعنى معظم القيم بالقرب من الوسط الحسابي والذيلين، أما في حالة ما يكون معامل التقلطح للبيانات الأصلية أقل من 3 يعنى ذلك أن المنحنى مفلطح بمعنى معظم القيم بعيدة عن الوسط والذيلين، أما في حالة ما يكون معامل التقلطح يساوي ثلاثة يكون المنحنى متوسط التقلطح (محمد، 2007م).

يبين الجدول (2) بعض المعايير الإحصائية للمساحة المنزرعة من المحاصيل تحت الدراسة، ويبين الجدول (3) بعض المعايير الإحصائية لإنتاجية الهكتار للمحاصيل تحت الدراسة، بينما الجدول (4) يوضح بعض المعايير الإحصائية لكمية الإنتاج للمحاصيل تحت الدراسة. ومن الجدول (2) نجد أن معامل الاختلاف للمساحة المنزرعة يتراوح ما بين 1.62% و 42.18%، بينما من الجدول (3)، نجد أنه يتراوح ما بين 2.97% إلى 16.87% لإنتاجية الهكتار للمحاصيل تحت الدراسة، والجدول (4) يبين قيمته التي تتراوح ما بين 3.78% إلى 35.07% لكمية الإنتاج للمحاصيل تحت الدراسة.

الجدول 2. بعض المعايير الإحصائية للمساحة المنزرعة من المحاصيل تحت الدراسة في مصر خلال الفترة 2006-2016.

معايير إحصائية	المحاصيل تحت الدراسة					
	الحبوب	الدرنات والجزور	قصب السكر	بنجر السكر	السهم	عباد الشمس
	(ألف هكتار/سنة)					
أقل قيمة	2883.6	120.15	133.02	78.29	25.10	6.38
أكبر قيمة	3224.4	209.46	140.73	235.09	42.30	17.14
المتوسط الحسابي	3073.8	167.62	136.83	160.66	31.77	10.01
الانحراف المعياري	117.57	26.14	2.22	54.78	5.39	4.22
الالتواء	-0.17	-0.25	0.07	-0.03	0.62	0.85
التقلطح	-1.43	-0.28	-0.14	-1.40	-0.14	-1.17
معامل الاختلاف (%)	3.82	15.59	1.62	34.10	16.97	42.18

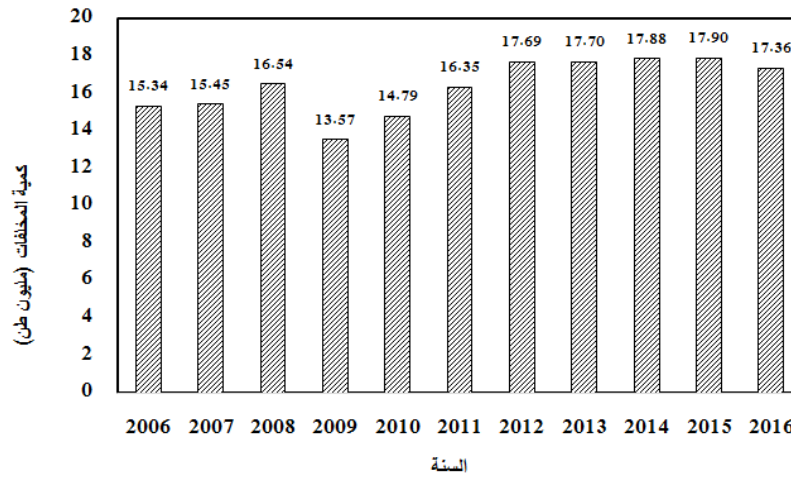
الجدول 3. بعض المعايير الإحصائية لإنتاجية المحاصيل تحت الدراسة في مصر خلال الفترة 2006-2016.

معايير إحصائية	المحاصيل تحت الدراسة					
	الحبوب	الدرنات والجذور	قصب السكر	بنجر السكر	السّمسم	عباد الشمس
	(كجم/هكتار)					
أقل قيمة	5739.0	12709.0	110487.4	47427.00	1207.00	2408.00
أكبر قيمة	7792.0	27822.0	121324.0	52337.00	1373.28	3376.36
المتوسط الحسابي	7265.4	25663.8	116405.2	49977.01	1309.02	2728.80
الانحراف المعياري	564.8	4330.1	3455.59	1913.74	47.53	333.15
الالتواء	-2.35	-3.22	0.22	-0.08	-0.94	0.88
التفطح	5.90	10.55	-0.53	-1.88	1.16	-0.54
معامل الاختلاف (%)	7.78	16.87	2.97	3.83	3.63	12.21

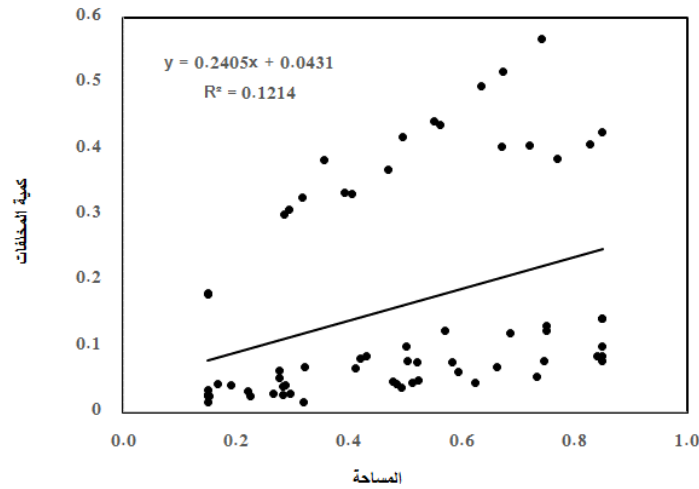
الجدول 4. بعض المعايير الإحصائية لكمية الإنتاج للمحاصيل تحت الدراسة في مصر خلال الفترة 2006-2016.

معايير إحصائية	المحاصيل تحت الدراسة					
	الحبوب	الدرنات والجذور	قصب السكر	بنجر السكر	السّمسم	عباد الشمس
	(ألف طن/سنة)					
أقل قيمة	17585	2662.03	14903.34	3905.18	32.74	18.93
أكبر قيمة	24034	5363.73	17014.68	11982.9	51.06	41.27
المتوسط الحسابي	22341	4252.54	15929.40	8051.26	41.45	26.41
الانحراف المعياري	2031.7	827.35	602.05	2823.23	6.06	8.88
الالتواء	-1.70	-0.68	0.38	-0.03	0.03	0.90
التفطح	2.25	-0.13	0.06	-1.56	-0.93	-1.06
معامل الاختلاف (%)	9.09	19.46	3.78	35.07	14.63	33.63

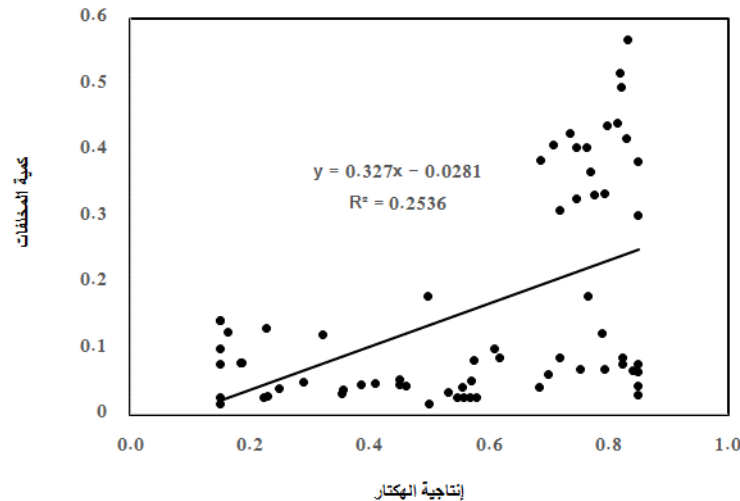
يوضح الشكل (5) توزيع كمية المخلفات الكلية تبعاً للسنة، ومن الشكل نجد أن أقل قيمة 13.57 مليون طن/سنة وأكبر قيمة 17.90 مليون طن/سنة، وتم حساب معامل الاختلاف بين هذه الكميات وكانت التفاوت بسيط بين كميات المخلفات ما بين سنة 2006م إلى سنة 2016م، حيث كان معامل الاختلاف حوالي 8.91%، مما يدل على أن نمط الإنتاج المزرعي خلال 11 سنة لم يتغير كثيراً في جمهورية مصر العربية في السنوات تحت الدراسة. ويبين الشكل (6) العلاقة بين المساحة المنزرعة وكمية المخلفات الزراعية الناتجة من المحاصيل تحت الدراسة بعد تعديل الأرقام ما بين 0.15 و 0.85، ومن الشكل نجد أن العلاقة الخطية بينهما ضعيفة من خلال قيمة معامل التحديد (R^2) والتي كانت 0.1214، وربما تكون العلاقة بينهما غير خطية، أما الشكل (7) فيوضح العلاقة بين إنتاجية الهكتار وكمية المخلفات الزراعية الناتجة من المحاصيل تحت الدراسة بعد تعديل الأرقام ما بين 0.15 و 0.85، ومن الشكل نجد أن العلاقة الخطية بينهما ضعيفة من خلال قيمة معامل التحديد (R^2) والتي كانت 0.2536، وربما تكون العلاقة بينهما غير خطية والذي يبين أهمية استخدام أدوات التفتيح عن البيانات، والتي يمكنها إيجاد العلاقات الغير خطية بين المتغيرات.



الشكل 5. توزيع كمية المخلفات الكلية تبعاً للسنة في مصر خلال الفترة 2006-2016.



الشكل 6. العلاقة بين المساحة المنزرعة وكمية المخلفات الناتجة من المحاصيل تحت الدراسة بعد تعديل الأرقام ما بين 0.15 و0.85 في مصر خلال الفترة 2006-2016.



الشكل 7. العلاقة بين إنتاجية الهكتار وكمية المخلفات الناتجة من المحاصيل تحت الدراسة بعد تعديل الأرقام ما بين 0.15 و0.85 في مصر خلال الفترة 2006-2016.

تقييم أدوات التنقيب عن البيانات للتنبؤ بكمية المخلفات الزراعية:

تم استخدام التطبيق الحاسوبي WEKA بعد ضبط القراءات وعمل ملف بيانات له الامتداد arff، ليتوافق مع التطبيق، وتم استخدام 83% من البيانات (55 قراءة) عشوائياً، كبيانات تدريب و17% (11 قراءة)، منها كبيانات اختبار للنموذج (11 قراءة)، وتم اختيار كل أدوات التنقيب عن البيانات التي تتعامل مع البيانات الكمية داخل WEKA لتنفيذ عملية الانحدار بين 7 مدخلات ومخرج واحد، دون تعديل لثوابت هذه الأدوات داخل التطبيق (default). ويبين الجدول (5) أدوات التنقيب عن البيانات المستخدمة ومفرداتها في تطبيق WEKA والزمن المستغرق لبناء النموذج. ويتبين من الجدول (5) أن الزمن المستغرق لبناء النماذج يتراوح ما بين 0 ثانية إلى 0.13 ثانية، وهذا يجب أخذه في الاعتبار عند اختيار الأداة المناسبة.

تم ترتيب الأدوات بحسب المعادلة (3)، وتبين أن أداة جدول القرار (DecisionTable) هي أفضل أداة للتنبؤ بكمية المخلفات الزراعية، كما هو مبين في الجدول (5) بحسب أن رتبته هي 1 بناءً على قيمة Rank score، وقيمة متوسط انحراف الخطأ (MBE) لها ومقداره -9.149 ألف طن/سنة كما يتبين من الجدول (5) أن قيم معامل التحديد لجميع الأدوات كان فوق 95%، وبحسب الإشارة لقيم متوسط انحراف الخطأ، نجد أن من أدوات التنقيب عن البيانات تتنبأ بقيم أقل من القيم الفعلية ومنها أدوات تتنبأ بقيم أكبر من القيم الفعلية. ومن فحص جدول (6) نجد أن المتوسط الحسابي لكمية المخلفات المتنبأ بها من 17 أداة للتنقيب عن البيانات متقارب، حيث أقل متوسط حسابي هو 2502.17 ألف طن/سنة وأكبر متوسط حسابي هو 3579.66 ألف طن، بينما المتوسط الحسابي للقيم الفعلية في بيانات الاختبار كان 3300.06 ألف طن/سنة.

الجدول 5. الزمن المستغرق لبناء النموذج (ثانية) وقيم معامل التحديد (R^2)، متوسط انحراف الخطأ (MBE)، ومتوسط الجذر التربيعي لمربع الخطأ (RMSE) وقيمة الإحصاء T ورتبة كل أداة لأدوات التنقيب عن البيانات.

أداة التنقيب عن البيانات	الزمن المستغرق لبناء النموذج (ثانية)	R^2	MBE	RMSE	T	Rank Score	الرتبة
Decision Table	0.01	0.991	-9.149	365.924	0.056	0.167	1
M5P	0.01	0.9925	19.322	343.18	0.126	0.235	2
Regression By Discretization	0.03	0.9759	26.542	560.358	0.106	0.281	3
M5Rules	0.05	0.9832	-32.144	402.588	0.179	0.314	4
Bagging	0.02	0.9933	-42.111	311.425	0.305	0.41	5
REPTree	0	0.9802	-64.128	450.005	0.322	0.48	6
Gaussian Processes (GP)	0.13	0.9945	96.723	559.262	0.393	0.595	7
Additive Regression	0.01	0.9933	-80.794	324.153	0.575	0.696	8
Kstar	0	0.9867	-93.216	386.543	0.556	0.701	9
Linear Regression (LR)	0.01	0.9848	158.672	686.723	0.531	0.794	10
SMOreg	0.02	0.9932	-102.178	310.205	0.78	0.9	11
Random Tree	0	0.9814	-174.194	492.959	0.845	1.043	12
IBK	0	0.982	-178.528	482.541	0.89	1.087	13
Random Forest	0.06	0.9896	-155.894	393.125	0.966	1.127	14
Multilayer Perceptron (MLP)	0.08	0.9917	-232.611	407.721	1.553	1.732	15
LWL	0	0.9967	259.454	350.583	2.461	2.663	16
Decision Stump	0.01	0.9591	663.422	951.725	2.174	2.819	17

الجدول 6. توزيع المتوسط الحسابي لكمية المخلفات المتنبأ بها من 17 أداة للتنقيب عن البيانات ومتوسط الكمية الفعلية داخل بيانات الاختبار

أداة التنقيب عن البيانات	المتوسط الحسابي لكمية المخلفات (ألف طن/سنة)
GP	3249.31
LR	3209.41
MLP	3579.66
SMOreg	3445.74
IBK	3369.17
Kstar	3308.95
LWL	3008.64
AdditiveRegression	3364.17
Bagging	3371.76
RegressionByDiscretization	3358.47
DecisionTable	3365.49
M5Rules	3221.70
DecisionStump	2502.17
M5P	3320.81
RandomForest	3405.33
RandomTree	3365.47
REPTree	3249.11
متوسط الكمية الفعلية	3300.06

الاستنتاجات:

بينت النتائج قدرة هذه الأدوات على التنبؤ بسبب العلاقة الغير خطية بين كمية المخلفات الزراعية والعوامل المؤثرة عليها، حيث تم الاعتماد على حساب كمية المخلفات بحسب الطريقة المتبعة في المنظمة العربية للتنمية الزراعية. وتبين أن طريقة جدول القرار (DecisionTable) هي أفضل أداة للتنبؤ بكمية المخلفات الزراعية من بين 17 أداة تم تقييمها، حيث كان قيمة متوسط انحراف الخطأ (MBE) لها 9.149- ألف طن/سنة وأن قيم معامل التحديد (R^2) لجميع الأدوات كان فوق 95%.

التوصيات:

توصي الدراسة باستخدام أدوات التنقيب عن البيانات في تحليل البيانات في مجالات الزراعة المختلفة لاستخلاص علاقات بين المتغيرات وتلخيصها إلى معلومات ذات فائدة. كما توصي الدراسة بضرورة الاهتمام الكافي بالتنبؤ بالمخلفات الزراعية من خلال استخدام النماذج الرياضية، حيث يمكن إدارة هذه المخلفات والاستفادة منها في إنتاج الطاقة أو السماد أو كعلف حيواني.

المراجع:

- إمام، دينا حسن، وهدى محمد الجنيهي وخيري حسن أبو السعود (2007). سلوك المرأة الريفية في التعامل مع المخلفات المزرعية والمنزلية بقرية دار السلام بمحافظة الفيوم. مجلة العلوم الزراعية، جامعة المنصورة. 32 (9): 7883-7893.
- الجناعي، أواب، والحسين الحداد، وعلى البار وعمار الزهاري (2011). استكشاف بعض الأنماط المؤثرة في الأداء الأكاديمي لطلاب جامعة العلوم والتكنولوجيا باستخدام تقنيات التنقيب في البيانات. مجلة العلوم والتكنولوجيا. 16(1): 22-38.
- حبيب، محمد حسب النبي وسعيد عباس محمد رشاد (2016). دراسة لمستوي معارف المرأة الريفية للمحافظة علي البيئة من التلوث في محافظة القليوبية. مجلة حوليات العلوم الزراعية بمشهر. 54 (1): 227-240.
- الحجار، صلاح محمود (2003م). السحابة الدخانية المشكلة. الأثر. الحل. الطبعة الأولى، دار الفكر العربي، القاهرة، جمهورية مصر العربية. 168 صفحة.
- الدوري، زكريا مطلق و داليا عبد الحسين أحمد (2007م). دور تنقيب البيانات Data Mining في زيادة أداء المنظمة" دراسة تحليلية في المصرف الصناعي. مجلة العلوم الاقتصادية والإدارية. 13(48): 40-62.
- سردينة، ابتسام عبد السلام عبد المنعم (2004م). دراسة السلوك البيئي لزوجات الخريجين والمنتفعين ببعض قرى منطقة البستان بمحافظة البحيرة. رسالة ماجستير غير منشورة، كلية الزراعة، جامعة الإسكندرية، جمهورية مصر العربية، 121 صفحة.
- سيد، أحمد فايز أحمد (2015) أدوات التنقيب عن البيانات مفتوحة المصدر دراسة تحليلية تقييمية. مجلة جامعة طيبة للآداب والعلوم الإنسانية. 5(10): 791-864.
- شاهين، هيثم، وسامية شيبان وسماح عباس (2017م). تطوير عملية إدارة النفايات الصلبة البلدية في مدينة اللاذقية باستخدام نظام System Dynamics. مجلة جامعة تشرين للبحوث والدراسات العلمية، سلسلة العلوم الهندسية. 39(2): 233-252.
- الشايب، آمال وسماح محمد (2011). احصاءات البيئة والطاقة في مصر. شوهده في 2020/6/6، في <https://unstats.un.org/unsd/energy/meetings/2013ees/2013negyp.pdf>. 21 صفحة.

- علي، هدى عبد الرحيم حسين (2018). استخدام تقنية التنقيب عن البيانات لتحليل المؤشرات المالية لعينة من المصارف الأهلية العراقية باعتماد خوارزمية CART. المجلة العراقية لتكنولوجيا المعلومات. 9 (2):32-54.
- محمد، أماني موسى (2007م) التحليل الإحصائي للبيانات. الطبعة الأولى مارس 2007 الناشر: مركز تطوير الدراسات العليا والبحوث في العلوم الهندسية كلية الهندسة، جامعة القاهرة رقم الإيداع بدار الكتب المصرية: 123/5673، 2007، صفحة.
- مكاوي، حنان فتحي ذكي محمد (2010). برنامج إرشادي منبثق من دراسة سلوك المرأة الريفية في الحفاظ على البيئة بمحافظة البحيرة. أطروحة دكتوراه غير منشورة، قسم الاقتصاد المنزلي، كلية الزراعة، جامعة الإسكندرية، جمهورية مصر العربية، 267 صفحة.
- المنظمة العربية للتنمية الزراعية (2006). دراسة تدوير المخلفات الزراعية للاستعمالات الصناعية والمنزلية في الدول العربية. الخرطوم، السودان.
- المنظمة العربية للتنمية الزراعية (2017). الكتاب السنوي للإحصاءات الزراعية العربية، جامعة الدول العربية، المجلد رقم 37.
- نصير، رحاب عبد الفتاح (2018). رؤية مستقبلية لإعادة تدوير المخلفات الزراعية ومدى تأثير تطبيقها على تصميم الأثاث والفراغ الداخلي. مجلة العمارة والفنون والعلوم الإنسانية. 3 (1): 215-231.
- Abbasi, M.; and A. El Hanandeh (2016). Forecasting municipal solid waste generation using artificial intelligence modelling approaches. *Waste Management*. 56:13-22.
- Abou El-Azayem, M.G.M.; and S.S. Abd El-Ghani (2010). Economic return of recycling the agricultural wastes in Egypt and Spain. *Journal of American Science*. 6:960-970.
- Abou Hussein, S.D.; and O.M. Sawan (2010). The utilization of agricultural waste as one of the environmental issues in Egypt (A case study). *Journal of Applied Sciences Research*. 6(8):1116-1124.
- Achinas, S.; G. Jan; and W. Euverink (2016). Theoretical analysis of biogas potential prediction from agricultural waste. *Resource-Efficient Technologies*. 3:143-147.
- Al-Subu, N.M. (2015). A mathematical model to predict the components' generation of solid waste and residents' concerns towards solid waste management facilities in Nablus and Jenin Districts. Unpublished Master Thesis, Water and Environmental Engineering, Faculty of Graduate Studies, Birzeit University, Palestine. https://fada.birzeit.edu/bitstream/20.500.11889/1758/1/thesis_14112015_103055.pdf, 93 pages.
- An, A.; S. Khan; and X. Huang (2003). Objective and subjective algorithms for grouping association rules. Conference: Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM 2003), 19-22 December 2003, Melbourne, Florida, US:1-4.
- Barnard, G.; and L. Kristoferson (1985). Agricultural residues as fuel in the third world, *Earth scan Energy Inf. Programme*, London, 178 pp.
- Hall, M.; E. Frank; G. Holmes; B. Pfahringer; P. Reutemann; and I.H. Witten (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, University of Waikato. 1:10-18.
- Heilala, T. (2018). Waste generation profiling by applying data-mining methods to finish community waste weight data. Master's Thesis, University of Eastern Finland, Department of Environmental and Biological Sciences. Finland, 61 pages.

- Jiang, P.; Y.V. Fan; J. Zhou; M. Zheng; X. Liu; and J.J. Klemeš (2020). Data-driven analytical framework for waste-dumping behaviour analysis to facilitate policy regulations. *Waste Management*.103:285-295.
- Kaewdiew, J.; R. Ramaraj; S. Koonaphapdeelert; and N. Dussadee (2019). Assessment of the biogas potential from agricultural waste in northern Thailand. *Maejo International Journal of Energy and Environmental Communication*.1:40-47.
- Kannangara, M.; R. Dua; L. Ahmadi; and F. Bensebaa (2018) .Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. *Waste Management*.74:3-15.
- Khan, M.M.; and M.J. Ahmad (2012). Estimation of global solar radiation using clear sky radiation in Yemen. *International Research Journal of Engineering Science, Technology and Innovation (IRJESTI)*.9: 228-237.
- Kodeeshwari, R.S.; and K.T. Ilakkiya (2017). Different types of data mining techniques used in agriculture - A survey. *International Journal of Advanced Engineering Research and Science (IJAERS)*. 6: 17-23.
- Kolebaje, O.T.; and L.O. Mustapha (2012). On the performance of some predictive models for global solar radiation estimate in tropical stations: Port Harcourt and Lokoja. *The African Review of Physics*.15:145-163.
- Kolebaje, O.T.; A. Ikusika; and P.Akinyemi (2016). Estimating solar radiation in Ikeja and Port Harcourt via correlation with relative humidity and temperature .*International Journal of Energy Production and Management*. 3:253-262.
- Kumar, A.; and S.R. Samadder (2017.) An empirical model for prediction of household solid waste generation rate – A case study of Dhanbad, India. *Waste Management*. 68:3-15.
- Loukil, F.; and L. Rouached (2020) Waste collection criticality index in African cities. *Waste Management*.103:187-197.
- Maddison, A.L.; A. Camargo-Rodriguez; I.M. Scott; C.M. Jones; D.M.O. Elias; S. Hawkins; A. Massey; J. Clifton-Brown; N.P. McNamara; I.S. Donnison and S.J. Purdy (2017). Predicting future biomass yield in *Miscanthus* using the carbohydrate metabolic profile as a biomarker. *Glob Change Biol Bioenergy*. 9(7):1264-1278.doi: 10.1111/gcbb.12418
- Ogolo, E.O. (2010). Evaluating the performance of some predictive models for estimating global solar radiation across the varying climatic conditions in Nigeria. *Pacific Journal of Science and Technology*. 1:60-72.
- Vesta Services (2000). Qnet2000 Shareware, Vesta Services, Inc., 1001 Green Bay Rd, STE 196, Winnetka, IL 60093.
- Yevich, R.; and J.A. Logan (2003). An assessment of biofuel use and burning of agricultural waste in the developing world. *Global Biogeochemical Cycles*. 4: 6-1-6-21.

Predicting the Quantity of Agricultural Wastes Using WEKA Application

Abdulwahed M. Aboukarima^{*(1,2)} Samy G. Hemeda⁽²⁾ Mohamed S. El-Marazky⁽²⁾ Riham El_Oliemy⁽²⁾ and Ibrahim S. Tabash⁽¹⁾

(1). Department of Agricultural Engineering, Faculty of Food Sciences, King Saud University, Riyadh, Saudi Arabia.

(2). Agricultural Engineering Research Institute, Agricultural Research Centre, Giza, Egypt

(*Corresponding author: Dr. Abdulwahed M. Aboukarima. E-Mail: aboukarima@gmail.com)

Received: 29/03/2020

Accepted: 06/06/2020

Abstract

In this study, a comparison was made between data mining tools in WEKA open source computer application to predict the amount of agricultural wastes in Egypt from the year 2006 to 2016. The amount of wastes was calculated according to the method presented in the Arab Organization for Agricultural Development (AOAD). The results showed that the lowest amount of wastes was 13.57 million tons/year and the largest amount was 17.90 million tons/year. The results revealed that the data mining tool called Decision Table was the best tool for predicting the amount of agricultural wastes. The predicted lowest and the largest means were 2502.17 and 3579.66 thousand tons/year, respectively, while the mean actual value in the test data set was 3300.06 thousand tons/year. The study recommended that adequate attention was directed to forecast agricultural wastes amounts, as these wastes can be managed and utilized in the production of energy, compost or animal feed.

Key words: Machine learning, Agricultural wastes, Modeling, Environment.