

التنبؤ باستخدام التنقيب في البيانات لتحسين عملية اتخاذ القرار

Predicting using data mining to enhance the decision-making process

العيفة أصالة^{*1}

¹ مخبر المغرب الكبير: الاقتصاد والمجتمع، جامعة عبد الحميد مهري قسنطينة 02،

assala.laifa@univ-constantine2.dz

تاريخ النشر: 2023/12/31

تاريخ القبول: 2023/12/31

تاريخ الاستلام: 2023/11/01

ملخص:

تهدف الدراسة الى استخدام الانحدار الخطي كتقنية من تقنيات التنقيب في البيانات للتنبؤ بالطاقة الانتاجية والتكاليف لمؤسسة اسمنت جزائرية، وذلك عن طريق استخدام برنامج التنقيب weka، وذلك لتمكين المؤسسة من وضع تحليل منطقي للوضعية المالية المستقبلية لها لتحسين جودة ومسار القرارات المتخذة.

حيث توصلت الدراسة الى جملة من النتائج تمثلت في الزامية قيام المنظمات بدراسات قائمة على طرق الذكاء الاصطناعي والرياضيات وغيرها لمعرفة الطريق الذي تسير عليه اذا كان فيه انحرافات ليتم معالجتها في الوقت المناسب وبأقل تكلفة ممكنة خاصة وأن الطرق التقليدية يتعذر عليها القيام بتحليل البيانات اذا كانت قاعدة البيانات ضخمة.

كلمات مفتاحية: التنقيب في البيانات، التنبؤ، اتخاذ القرار، الانحدار الخطي.

تصنيفات JEL : C22,C25,C53.

Abstract:

The study aims to use linear regression as a technique among data mining techniques to predict the productive energy and costs for an Algerian cement company. This will be achieved through the utilization of the Weka data mining software. The objective is to enable the company to conduct a logical analysis of its future financial situation, thereby enhancing the quality and direction of the decisions made.

The study concluded that organizations are compelled to conduct studies based on artificial intelligence, mathematical methods, and others to understand the path they are taking. This is particularly important for identifying deviations that need to be addressed in a timely manner and at minimal costs. Traditional methods, especially when dealing with massive databases, face challenges in analyzing data.

Keywords: Data mining, forecasting, decision making, linear regression.

JEL Classification Codes: C22, C25, C5.

1. مقدمة:

تعتبر عملية اتخاذ القرار العامل الأساسي الفاصل بين نجاح المنظمة من عدمه، خاصة وأن بيئة العمل الاقتصادية اليوم تتميز بالانفتاح العالمي بين الأسواق، مما يجعلها بحاجة الى الدمج بين كل من الأساليب النوعية القائمة على الخبرة والتجربة والأساليب العلمية الرياضية التي تعطي بدورها تفسيرات منطقية تساهم بشكل كبير في تحسين جودة القرارات المتخذة وفي الوقت المناسب، ولعل من أبرز التقنيات المستعملة اليوم لدى العديد من المحللين الماليين والاقتصاديين وعلماء البيانات ما يسمى بالتنقيب في البيانات.

وللخروج بتصور واضح عن اشكالية الدراسة قمنا بصياغتها في التساؤل الرئيسي التالي:

كيف تساهم تقنيات التنقيب في البيانات في تحسين جودة القرارات المتخذة؟

والتي انبثقت عنها مجموعة من الفرضيات الأساسية المتمثلة في:

- تساعد تقنيات التنقيب في البيانات بشكل كبير في تحسين جودة القرارات المتخذة;
- الاعتماد على الانحدار الخطي كتقنية تنبؤي لا يساعد في الكشف عن الوضعية المالية المستقبلية للمنظمة .

2.1. أهداف الدراسة: تمثلت الغاية الأساسية من البحث في النقاط التالية:

- ✓ التطرق للإطار المفاهيمي للتنقيب في البيانات ;
 - ✓ تبين خطوات استخدام برنامج weka عند القيام بالتنبؤ لتسهيل استعمله من قبل الباحثين الآخرين;
 - ✓ إبراز الدور الكبير لتقنية التنبؤ باستخدام الانحدار الخطي في ترشيد القرارات المتخذة ;
- 3.1. منهجية الدراسة:** تهدف دراسة اشكالية الدراسة والاجابة عنها قمنا بالاعتماد على المنهج الوصفي عند التطرق للأدبيات النظرية المتعلقة بالتنقيب في البيانات، وكذا المنهج القياسي عند القيام بالتنبؤ بالطاقة الانتاجية والتكاليف المترتبة عنها في المدة الزمنية من 2022 الى غاية 2026 باستخدام برنامج Weka.

2. الإطار النظري للدراسة:

سنتناول في هذا الجزء من الدراسة الإطار النظري لكل من ادوات التنقيب في البيانات ومساهماتها في عملية اتخاذ القرار التي بدورها تساعد في تحسين جودة أداء المنظمات.

1.2. تعريف التنقيب في البيانات:

تعددت المراجع التي تناولت تعريف التنقيب نظرا للاهتمام الكبير بهذا المجال من قبل العديد من الباحثين خاصة في الاونة الأخيرة والتي يمكن ربطها بصفة كبيرة بعصر المعلوماتية حيث ينقسم التنقيب بدوره الى ثلاثة أنواع رئيسية تتمثل في: (كادي و خديم، 2015، صفحة 147)

- ✓ التنقيب في البيانات Data Mining: يعمل على معالجة البيانات المنظمة في قاعدة بيانات كبيرة ;
- ✓ تنقيب النصوص Text Mining: خاص بالبيانات والنصوص الغير مهيكلة;

✓ تنقيب الويب Web Mining: يتميز بكونه يعمل على معالجة البيانات التي تقع بين البيانات شبه الهيكلية والبيانات الغير مهيكلة، ويعتبر من أصعب البيانات التي يعمل التنقيب على معالجتها وإدارتها، بسبب الكم الضخم من البيانات الغير مرئية والغير مترابطة المتواجدة على الأنترنت.

أما دراستنا فقد اقتصرنا على تناول التنقيب في البيانات باعتباره أداة حديثة يعمل الاقتصاديين على دمج كل من تقنياتها وتقنيات الاقتصاد القياسي في تحليل البيانات.

حيث عرف (PANG & Authers, 2006, pp. 01-03) التنقيب في البيانات على أنه عملية دقيقة الية تتميز بالتلقائية في الأداء تعمل على استخراج المعلومات المفيدة من مستودع البيانات الكبيرة، لاستنتاج معارف لم تكن معروفة مسبقا، وذلك عن طريق مزج العديد من طرق التحليل التقليدية، الذكية والخوارزميات المعقدة لتحليل واستكشاف أنواع جديدة من البيانات.

أما (وحيد و عبد الواحد ، 2021، الصفحات 340-341) فقد رأى أن التنقيب في البيانات عبارة عن أسلوب أو عملية للبحث في كميات كبيرة من البيانات سواء كانت قاعدة بيانات أو مستودع بيانات أو عبارة عن بيانات ضخمة بهدف الحصول على النقاط التالية:

✓ الحصول على معلومات محددة ومستهدفة ;

✓ ايجاد علاقات منطقية من هذه البيانات;

✓ اكتشاف النماذج وكذا العلاقات المخبئة داخل قاعدة البيانات ;

✓ المساعدة في التنبؤ;

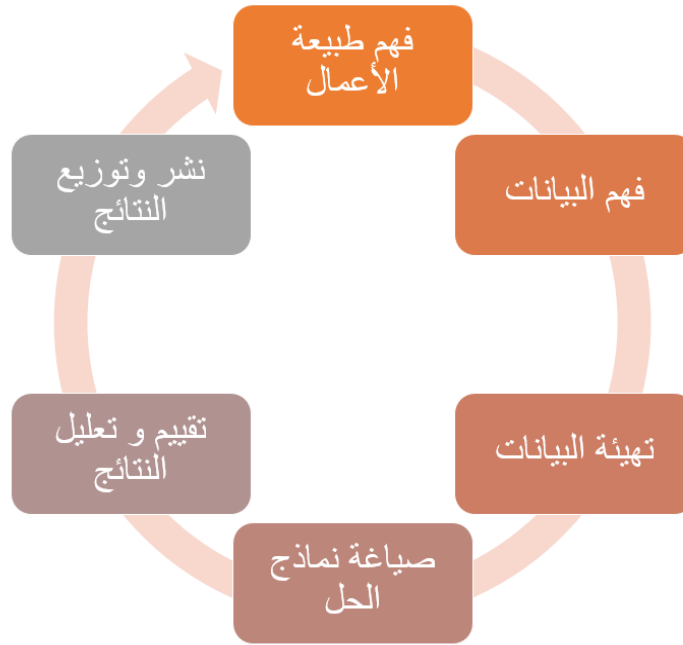
✓ استخلاص معلومات جديدة ومفيدة وذلك في الوقت المناسب ;

✓ استخلاص معلومات تساعد في عملية اتخاذ القرار.

وذلك عن طريق تحليل وتلخيص البيانات اعتمادا على تطبيقات الذكاء الاصطناعي والأساليب الاحصائية وكذا مختلف تقنيات تقانة المعلومات المعروفة بالتنقيب في البيانات.

ومن التعاريف السابقة نستنتج أن التنقيب في البيانات عبارة عن الية تعمل على استخراج معارف من كمية ضخمة من البيانات وذلك عن طريق تحليل، تفسير، استخراج العلاقات والأنماط الغير معروفة سابقا وتقديم تنبؤات تساعد في عملية صنع القرار باستخدام العديد من التقنيات والخوارزميات القائمة على الذكاء الاصطناعي والتعلم الآلي وكذا الطرق الاحصائية والرياضية، وذلك عن طريق مجموعة من المراحل المتسلسلة والتي يوضحها الشكل التالي:

الشكل رقم(01): دورة حياة التنقيب في البيانات



المصدر: من اعداد الباحثة بالاعتماد على (أحمد و محمد ، 2020 ، صفحة 12)

حيث يمكن شرح مختلف المراحل بشكل موجز كما يلي: (أحمد و محمد ، 2020 ، الصفحات 13-14)

- ✓ فهم طبيعة الأعمال: تتمثل تحديد الهدف الرئيسي من عملية التنقيب ;
- ✓ فهم البيانات: عن طريق فهم طبيعة البيانات عن طريق تجميعها من مختلف مصادرها ثم فرزها والتحقق من جودتها;
- ✓ تهيئة البيانات: عن طريق اختيار وترتيب حقول البيانات كما تتطلب عملية التنقيب في البيانات
- ✓ صياغة نماذج الحل: تحتاج هذه العملية المختصين في تحليل البيانات والتنقيب من أجل تحديد الخطأ والصواب لاختيار أفضل نموذج;
- ✓ تقييم وتعليق النتائج: التحقق من دقة النموذج عن طريق اجراء مقارنات بين النتائج المتوقعة والنتائج الفعلية;
- ✓ نشر وتوزيع النتائج: وذلك عن طريق نشر النموذج داخل المنظمة للمساعدة في عملية صنع القرار.

2.2. أنواع تقنيات التنقيب في البيانات:

وتنقسم تقنيات التنقيب في البيانات الى صنفين رئيسيين تتمثلان في :

1.2.2. التنقيب الوصفي: والذي يعمل على وضع تحليلات وتقديم وصف موجز ومفيد للبيانات قيد الدراسة ;

2.2.2. التنقيب التنبؤي: يشمل الخوارزميات التي تعمل على وضع توقعات مستقبلية لسلوك الظاهرة محل الدراسة اعتمادا على سلوكها في الماضي.

تعددت التقنيات المستعملة في التنقيب الوصفي نذكر منها التجميع، قواعد الارتباط، التلخيص وغيرها أما بالنسبة للتنقيب التنبؤي فيشمل كل من تقنية التصنيف، الانحدار وتحليل السلاسل الزمنية... (Atre, p. 64) حيث تركز دراستنا على تقنية الانحدار باعتباره أهم وأشهر تقنية لدى الاقتصاديين في فهم السلوك المستقبلي لدى المتغير قيد الدراسة عن طريق اكتشاف العلاقة بين المتغيرات في مجالات عديدة نذكر منها: (Harsh, 2022)

- ✓ التنبؤ أو التنبؤ المالي ;
- ✓ معايرة البيانات الإحصائية ;
- ✓ تقدير انتاج المحاصيل ;
- ✓ تخطيط الأعمال والتسويق.

ومن بين أنواعه ما يلي:

- **الانحدار الخطي البسيط:** من أبسط نماذج الانحدار والذي يعبر عن العلاقة بين متغير تابع واحد ومتغير مستقل واحد بحيث تكون معادلته كالتالي:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$$

بحيث:

Y_i : المتغير التابع;

β_0 : قيمة ثابتة تعبر عن قيمة Y عندما تكون قيمة x_{i1} تساوي الصفر;

β_1 : معامل الانحدار للمتغير المستقل;

x_{i1} : المتغير المستقل;

ϵ_i : الخطأ العشوائي.

- **الانحدار الخطي المتعدد:** عندما تكون العلاقة بين متغير تابع واحد وعدة متغيرات كمية "أكثر من متغير واحد" بواسطة المعادلة التالية:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \epsilon_i$$

بحيث:

β_0 : قيمة ثابتة تعبر عن قيمة Y عندما تكون قيمة $x_{i1}; x_{i2}$ تساوي الصفر;

$(x_{i1}; x_{i2})$: متغيرات مستقلة;

$(\beta_1; \beta_2)$: معاملات الانحدار للمتغيرات المستقلة. (سوسن ، 2013 ، صفحة 178)

- **الانحدار اللوجستي:** ويطلق عليه أيضا بالنموذج اللوجستي، حيث يقوم بتحليل العلاقة بين مجموعة من المتغيرات المستقلة ومتغير تابع يكون اما ترتيبيا أو تصنيفيا وذلك عن طريق تقدير احتمال وقوع حدث من عدمه عن طريق تركيب ما يسمى بالمنحنى اللوجستي. (محمد أمين و محمد ، 2017 ، صفحة 125)

3. الجانب التطبيقي:

1.3. تشخيص الطاقة الانتاجية بالشركة: تعتبر مؤسسة جيكا فرع قسنطينة من المؤسسات المتخصصة والرائدة في صناعة الاسمنت بشرق الجزائر، وفي هذا الجزء سنحاول تشخيص الطاقة الانتاجية وتكاليفها التي يوضحها لنا الجدول أسفله.

الجدول رقم(01): تطور كميات وتكاليف انتاج الاسمنت في الشركة محل الدراسة خلال الفترة

2021-2015

السنة	2015	2016	2017	2018	2019	2020	2021
الكمية "طن"	1 202 043	1 153 240	1 140 014	1 006 118	1 006 118	822 335	695 442
تكلفة الوحدة السنوية "د. ج"	3 605	3 106	3 608	3 128	3 642	4 194	4 848

المصدر: وثائق المحاسبة التشغيلية للشركة من 2015- 2021

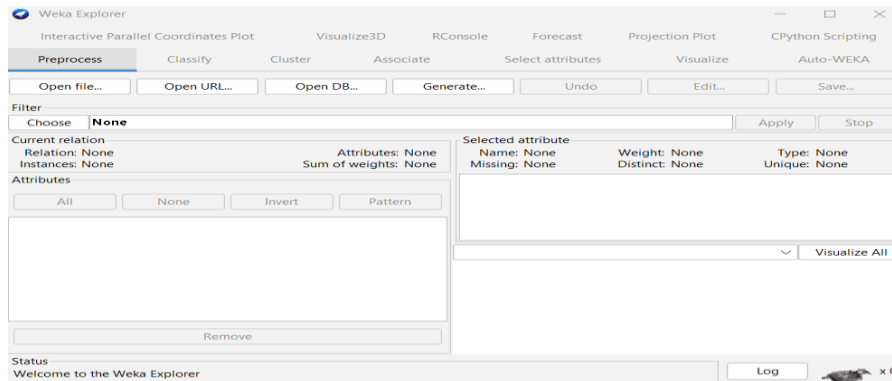
من خلال الجدول المرافق له نلاحظ ان كمية انتاج الاسمنت في السنوات السبع تتناقص تدريجيا من 1202043 طن الى 695 442 طن في سنة 2021، أما التكاليف فنلاحظ تذبذب من سنة 2015 الى سنة 2018 وفي سنة 2019 استمرت التكاليف في الارتفاع حتى 2021.

حيث يرجع ارتفاع التكاليف ابتداء من 2019 الى الأوضاع الاقتصادية التي شهدها الاقتصاد الجزائري بشكل خاص والعالم بشكل عام نتيجة التعرض لفيروس كوفيد 19 والذي بدوره أثر سلبا على جل القطاعات خاصة التجارية منها وهذا ما يؤكد التناقص التدريجي في كمية انتاج الاسمنت بسبب نقص الطلب لشلل قطاعات البناء بسبب الحجر الصحي وغيره من الأسباب التي أدت الى تدهور هذا القطاع بشكل نسبي مقارنة بالقطاعات الأخرى.

2.3. تطبيق أداة weka3.8.6: تعتبر هذه الاداة من ادوات التنقيب في البيانات تم استخدامه لسهولة ومجانيته، إن استخدام هذه الاداة في التنبؤ بالطاقة الانتاجية لمؤسسة الاسمنت كان باتباع الخطوات التالية:

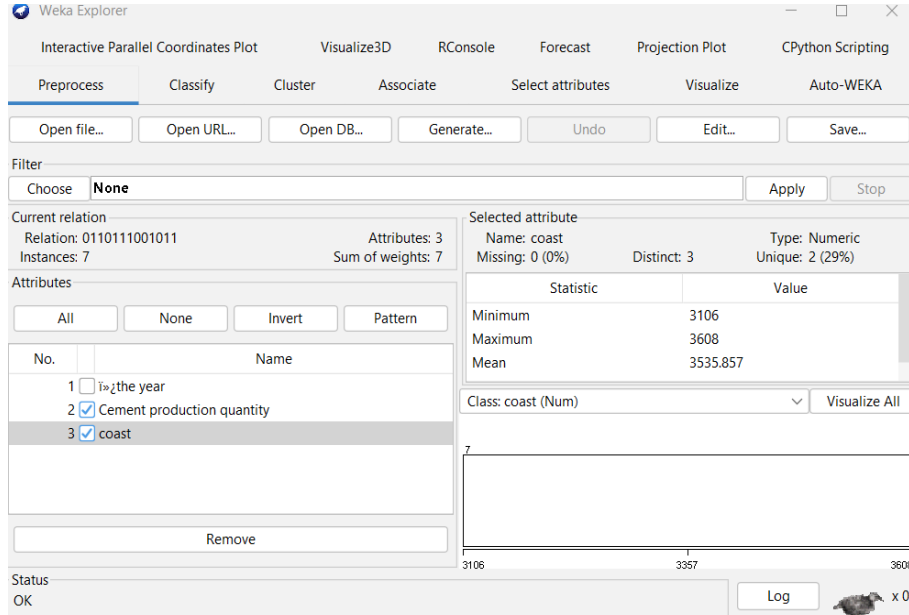
-يجب تحويل من Excel الى CSV ومن ثم تحميله في WEKA بالنقر على open file ...

الشكل رقم(02): واجهة برنامج Weka



المصدر: مخرجات برنامج Weka

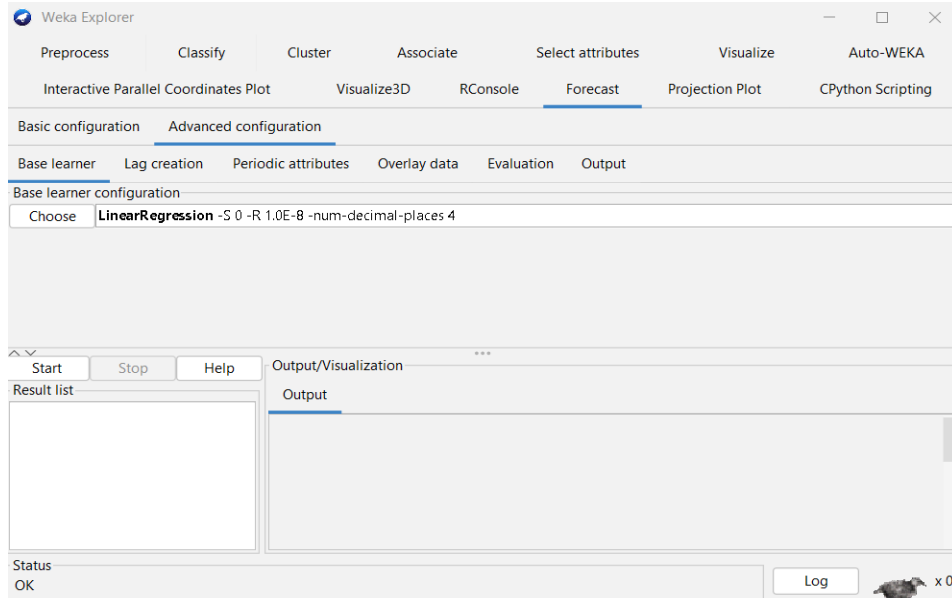
الشكل رقم(03): ادخال البيانات الى البرنامج



المصدر: مخرجات برنامج Weka

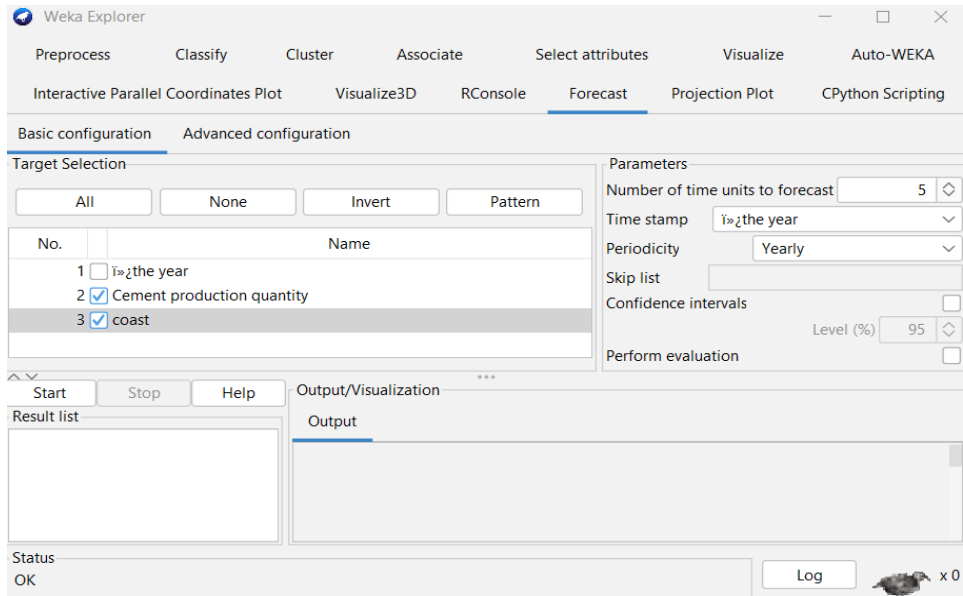
-النقر على forecast للتنبؤ بالمتغيرات التابعة و التي يتم تحديدها دون السنوات ثم تحديد عدد السنوات المراد التنبؤ بها المتمثلة في 5 سنوات.

الشكل رقم(04): خطوات عملية التنبؤ وفق برنامج Weka



المصدر: مخرجات برنامج Weka

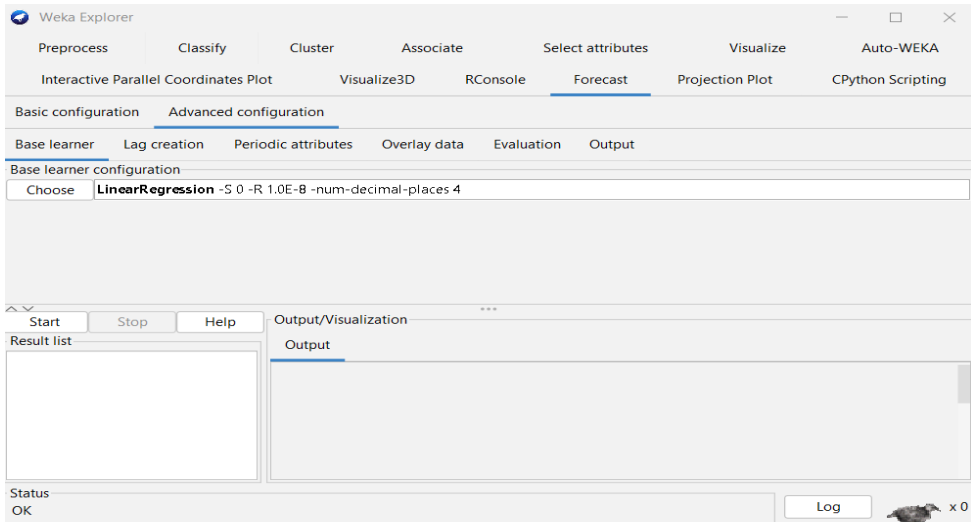
الشكل رقم(05): خطوات عملية التنبؤ وفق برنامج Weka



المصدر: مخرجات برنامج Weka

-اختيار Linear Regression باعتباره وسيلة تستخدم في التنبؤ بقيمة المتغير بناء على معرفة قيم متغير أو أكثر.

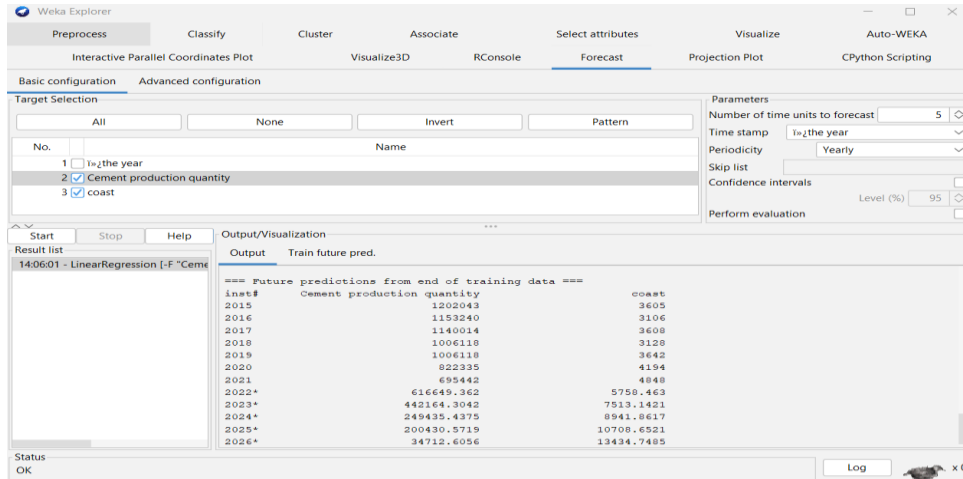
الشكل رقم(06): اختيار الانحدار الخطي وفق برنامج Weka



المصدر: مخرجات برنامج Weka

-بمجرد النقر على Start يتم اظهار نتائج التنبؤ بكل من كمية انتاج الاسمنت و التكاليف

الشكل رقم (07): نتائج عملية التنبؤ



المصدر: مخرجات برنامج Weka

والجدول الموالي يبين لنا النتائج حسب أداة weka :

الجدول رقم (02) : نتائج التنبؤ خلال الفترة 2026-2022 باستخدام برنامج weka

السنة	2022	2023	2024	2025	2026
الكمية المتوقعة "الطن"	616 649.36	442 164.30	249 435.43	200 430.57	34 712.60
تكلفة الوحدة المتوقعة "د. ج"	5 758.46	7 513.14	8 941.86	10 708.65	13 434.74

المصدر: من اعداد الباحثة بالاعتماد على مخرجات برنامج weka

حيث نلاحظ من الجدول أعلاه أن كميات الانتاج تستمر في الانخفاض في المقابل أن تكاليف الوحدة الواحدة تستمر بالارتفاع بشكل كبير، وبالتالي يعتبر مؤشر سلبى للشركة قد يؤدي الى وقوعها في الفشل المالي، لذا يجب على الشركة القيام باستراتيجيات أخرى تمكنها من خفض تكاليفها للقدرة على وضع أسعار وكميات تتناسب مع احتياجات المقاولين والعملاء خاصة وأن القطاع يشهد تنافس شديد في انتاج مادة الاسمنت، وبالتالي الزامية اتخاذ قرارات استراتيجية تمكنها من وضع خطة لمعالجة الانحرافات التي قد تتعرض لها الشركة مستقبلا والذي اذا استمرت دون معالجة قد تؤدي الى التصفية.

4.خاتمة: من خلال الدراسة النظرية والتطبيقية الموجزة توصلنا الى النتائج والتوصيات المتمثلة في :

1.4.النتائج:

-تساعد تقنيات التنقيب في البيانات بشكل كبير في تحليل البيانات عن طريق مهمتين رئيسيتين تتمثلان في الوصف والتنبؤ، وذلك بهدف تحسين جودة القرارات المتخذة ووضع رؤى مستقبلية لوضعية المنظمة وهذا ما يثبت صحة الفرضية الأولى " تساعد تقنيات التنقيب في البيانات بشكل كبير في تحسين جودة القرارات المتخذة";

-الزامية قيام المنظمات بدراسات قائمة على طرق الذكاء الاصطناعي والرياضيات وغيرها لمعرفة الطريق الذي تسير عليه اذا كان فيه انحرافات ليتم معالجتها في الوقت المناسب وبأقل تكلفة ممكنة خاصة وأن الطرق التقليدية يتعذر عليها القيام بتحليل البيانات اذا كانت قاعدة البيانات ضخمة ;

-الاعتماد على الانحدار الخطي كتنبؤ يساعده في الكشف عن الوضعية المالية المستقبلية للمنظمة وبالتالي صنع القرار المناسب، وهذا ما ينفي الفرضية الثانية التي تنص على " الاعتماد على الانحدار الخطي كتنبؤ يساعده في الكشف عن الوضعية المالية المستقبلية للمنظمة" .

-القدرة على تقليل الأخطاء في اتخاذ القرارات وذلك لسهولة استخدام بعض البرامج من قبل الاقتصاديين اذ أن الأمر لا يتطلب بالضرورة فهم كيفية النمذجة أو كيفية عمل التقنيات والخوارزميات بل كيفية تطبيق البرنامج وتحليل النتائج فقط ;

-تطبيق تقنيات الانحدار في التنقيب في البيانات يساعد المؤسسات في تحليل العلاقات وتقديم تنبؤ دقيق لمختلف المؤشرات ذات العلاقة بالظاهرة قيد الدراسة لاتخاذ قرارات تحسينية استراتيجية للؤسسة.

2.4.التوصيات:

-القيام بعدة ملتقيات ودورات تكوينية تخص كيفية تطبيق مختلف النماذج والأساليب الكمية من خلال الحواسيب الالية لأن الاشكالية اليوم لدى معظم الاقتصاديين تكمن في كيفية تطبيق هذه الخوارزميات والتقنيات القائمة على الذكاء الاصطناعي حيث وأن معظم برامج تحليل البيانات الرائدة مثل R ، Matlab ، Python تتطلب بشكل كبير الاختصاص في مجال الاعلام الالي، خاصة وأن مجال الاقتصاد والمالية اليوم يتطلب بشكل كبير الدمج مع تخصص تحليل البيانات باعتباره علم قائم بذاته وتخصص الاعلام الالي.

5.قائمة المراجع:

- Atre, S. (n.d.). Defining Today's Data Mining Executive Update. Business Intelligence Advisory Service, 01(04).
- Harsh, V. (2022, 05 27). Regression Method in Data Mining Simplified 101. Retrieved 08 19, 2022, from HEVO: <https://hevodata.com/learn/regression-method-in-data-mining/>
- PANG, N., & Authers . (2006). Introduction to Data Mining . united States of America: Pearson Education, Inc.

إبراهيم سوسن . (2013). التنبؤ بالتحصيل الدراسي لطلاب الجامعة من خلال بعض المتغيرات المعرفية والديموجرافية باستخدام كل من الانحدار الخطي المتعدد والانحدار اللوغاريتمي الثنائي. مجلة الدراسات التربوية والانسانية، 05(04).

- دعش محمد أمين ، وساري محمد . (2017). نموذج الانحدار اللوجستي: مفهومه، خصائصه، تطبيقاته "مع مثال تطبيقي للانحدار اللوجستي ثنائي الحدين على برنامج "spss مجلة السراج في التربية وقضايا المجتمع(01).
- زين الدين كادي ، وخديجة خديم. (2015). التنقيب المعلوماتي ودوره في تحليل احتياجات مستعملي المكتبات ومراكز المعلومات. مجلة العلوم الانسانية.
- عبد المطلب أحمد ، وحسن محمد . (2020). تنقيب البيانات التعليمية باستخدام خوارزمية (Apriori لتقييم أداء الطالب وتحسين الانتاج الأكاديمي. مجلة النيل الأبيض للدراسات والبحوث.
- محمود وحيد ، وغازي عبد الواحد . (2021). متطلبات تطبيق التنقيب في البيانات (minig Data في التدقيق تحديد مخاطر التدقيق نموذجاً. مجلة البحوث في العلوم المالية والمحاسبية، 06(01).