Syrian Arab Republic

Ministry of Higher Education & Scientific researches

Syrian Virtual University

الجمهورية العربية السورية

وزارة التعليم العالي والبحث العلمي

الجامعة الافتراضية السورية

نظام التنبؤ عن واردات نهر الفرات باستخدام تقنيات تعلم الآلة

Prediction system for the Incoming quantities of the Euphrates River using machine learning Techniques

بحث مقدم لنيل درجة الماجستير في تقانات الوب

إعداد الطالب:

مريم علي غزال Mariam_100528

إشراف:

الدكتور باسل الخطيب

2022

قرار تشكيل لجنة المناقشة

Syrian Arab Republic

Ministry of Higher Education and Scientific Research

Syrian Virtual University

الجمهورية العربية السورية

وزارة التعليم العالي والبحث العلمي

الجامعة الافتراضية السورية

قرار رقم / ١٥٢ / ٤٠

رئيس الجامعة الافتراضية السورية
بناءً على أحكام المرسوم التشريعي رقم 25 لعام 2002 وتعديلاته
وعلى مقترح مجلس الشؤون العلمية بجلسته رقم 366 المنعقدة بتاريخ 2022/03/01
وعلى إقرار مجلس الجامعة بجلسته رقم 358 المنعقدة بتاريخ 2022/03/01
يقرر ما يلي:

المادة الأولى: تُشكل لجنة الحكم على رسالة الماجستير التالية في برنامج تقانات الوب MWT، للطالبة مريم غزال (ID:100528)، بعنوان:

نظام التنبؤ عن واردات نهر الفرات باستخدام تقنيات تعلم الآلة

والمؤلفة من:

د. باسل الخطيب- جامعة دمشق- الاختصاص: الذكاء الصنعي	رئيساً ومشرفاً
د. محمد مازن محمد مصطفى- الجامعة الدولية للعلوم والتكنولوجيا- الاختصاص: برمجيات	عضواً
د. ماجدة بكور- الجامعة السورية الخاصة- الاختصاص: الذكاء الصنعي	عضواً

المادة الثانية: يبلغ هذا القرار من يلزم لتنفيذه.

دمشق في ١٣ / ٠٤ /2022

رئيس مجلس الجامعة الافتراضية السورية

الدكتور خليل عجمي

يوزع إلى
السيد الدكتور رئيس الجامعة
السيد الدكتور نائب رئيس الجامعة للشؤون الإدارية وشؤون الطلاب
مديرية شؤون الطلاب - الامتحانات - منسق البرنامج – مدير البرنامج – أمانة السر

الشكر والتقدير

الشكر العميق لأستاذي المشرف الدكتور باسل الخطيب، الذي قدم لي نموذجاً رائعاً في الإشراف والدفع الإيجابي، ولم يدخر أي جهد للإنجاز هذا البحث على أفضل وجه ممكن..... أدين له بالفضل.

وأتوجه بالشكر لأعضاء الهيئة التدريسية في الجامعة الافتراضية السورية لما يقدمونه من علم وجهد.

كما أتقدم بالشكر من السادة أعضاء لجنة المناقشة مسبقاً على تفضلهم لاستقراء محتويات هذا البحث وإثرائه.

الإهداء

إلى الذي أوليه حبي والذي أضاء سناه دربي.

إلى الظل الذي آوي إليه في كل حين.

إلى من زرع في قلبي أسمى معاني التفاؤل والصبر.

تتسابق الكلمات وتتزاحم العبارات لتنظم عقداً لا يستحقه إلا هو.

ملخص البحث

تكثر التحديات التي تواجه قطاع المياه والتي تؤثر سلباً على الامن المائي، والذي بدوره يؤثر على الأمن الغذائي، ومن بين التحديات الرئيسية عدم القدرة على تنفيذ خطط الإدارة المتكاملة للموارد المائية التي تشمل تأمين حصص مائية، وأثر تغير المناخ، ولمواجهة هذه التحديات يقدم تقرير (المياه والتنمية الثامن – أهداف التنمية المستدامة المتعلقة بالمياه في المنطقة العربية) المقدم من الاسكوا ESCWA ، تقييم الوضع المائي استناداً إلى التنبؤ على المديين القصير والمتوسط بالتغيرات في الطلب والامداد للتصدي للنقص في المياه في المستقبل .

نقوم في هذه الدراسة على التنبؤ بالواردات المائية لنهر الفرات باستخدام خوارزميات التصنيف. وبالاعتماد على نظام يقوم بتحليل الوارد المتوقع من دولة المنبع باستخدام عدة عوامل: الهطولات المطرية والثلجية – درجات الحرارة – سعر النفط العالمي – الموسم.

تُبرز الدراسة دور التقنيات والأساليب الحديثة وإمكانية تطبيقها في إدارة وتخطيط الموارد المائية بشكل عام وفي منطقة الدراسة بشكل خاص (نهر الفرات)، وتقدم قيمة مضافة لإدارة الموارد المائية مما ينعكس ايجاباً من الناحية الاجتماعية والاقتصادية، يزيد من كفاءة التخطيط المائي على مياه الفرات ويحقق الاستفادة القصوى في المشاريع المقامة عليه.

Abstract

There are many challenges facing the water sector, which negatively affect water security, and that's affect food security, among the main challenges is the inability to implement integrated water resource management plans, which include securing water quotas and the impact of climate change.

To meet these challenges, an ESCWA report (Water and Development the Eighth - Sustainable Development Goals Related to Water in the Arab Region), presents assessing the water situation based on short and medium-term forecasts of changes in demand and supply to address water shortages in the future.

In this study, we predict the water supply of the Euphrates River using classification algorithms. And by using a system that analyzes the expected supply from the upstream country relaying on several factors: rain and snow precipitation - temperatures - the world oil price - the season.

The study highlights the role of modern techniques and methods and the possibility of their application to the management and planning of water resources in general and in this study area in particular (Euphrates River). which provides an added value to the management of water resources, and it will positively reflected in social and economic terms, and increases the efficiency of water planning on the Euphrates waters to achieve maximum benefit in the projects built on the river.

محتويات الأطروحة

قرار تشكيل لجنة المناقشة 1

الشكر والتقدير 2

الإهداء 3

ملخص البحث 4

محتويات الأطروحة 6

فهرس الأشكال 10

فهرس الجداول 12

المصطلحات 14

مصطلحات فنية 14

مصطلحات تقنية 18

القسم النظري 25

الفصل الأول : الاطار العام للبحث 26

1.1 المقدمة 26

2.1 المشكلة العلمية ومبررات مشروع البحث 27

3.1 الهدف من البحث 27

4.1 أهمية البحث 28

5.1 النتائج التطبيقية المتوقعة من البحث والجهات المستفيدة منه 28

6.1 الجهات المستفيدة 28

7.1 الدراسات السابقة 29

1.7.1 الدراسة الأولى 29

2.7.1 الدراسة الثانية 30

3.7.1 الدراسة الثالثة 31

4.7.1 الدراسة الرابعة 32

5.7.1 الدراسة الخامسة 33

6.7.1 الدراسة السادسة 34

7.7.1 الدراسة السابعة 34

8.7.1 الدراسة الثامنة 35

9.7.1 الدراسة التاسعة 36

10.7.1 الدراسة العاشرة 37

11.7.1 الدراسة الحادية عشر 37

الفصل الثاني : المياه ونهر الفرات 39

1.2 مقدمة 39

2.2 الوضع المائي في سوريا 40

1.2.2 التحديات الرئيسية في قطاع المياه في الجمهورية العربية السورية 40

3.2 أهمية نهر الفرات 41

الفصل الثالث: الذكاء الصنعي..... 44

1.3 تطور البيانات 44

2.3 التنقيب البيانات 45

1.2.3 إجرائية التنقيب في البيانات..... 47

2.2.3 محاور التنقيب في البيانات Data Mining Tasks 48

1.2.2.3 *التنقيب في قواعد الاقتران*..... 48

2.2.2.3 *التصنيف* 49

3.2.2.3 *العنقدة*..... 50

3.3 خوارزميات التصنيف والتنبؤ 50

1.3.3 أنواع خوارزميات التصنيف 51

2.3.3 مصفوفة الارتباك 54

3.3.3 مثال عملي لطريقة التصنيف باستخدام نظرية الاحتمالات 56

القسم العملي 60

الفصل الرابع: الحل المقترح 61

1.4 جمع ومعالجة البيانات 61

2.4 مجموعة البيانات النهائية 64

3.4 الخوارزميات المقترحة 65

4.4 التنفيذ البرمجي 66

2.4.4 الوظائف البرمجية 67

الفصل الخامس: دليل الاستخدام 72

1.5 الواجهة الرئيسية 72

2.5 واجهة عن التطبيق 73

3.5 واجهة تحليل النتائج 74

4.5 واجهة التنبؤ 77

5.5 أمثلة عملية لاستخدام التطبيق 78

الفصل السادس: النتائج والمقارنات 83

1.5 تحليل المدخلات بيانياً 83

2.5 اختيار الخوارزميات المناسبة 86

3.5 التدريب 87

4.5 مناقشة النتائج 97

الفصل السادس: الخاتمة والأفاق المستقبلية 98

1.6 الخاتمة 98

2.6 الآفاق المستقبلية 99

المراجع 100

فهرس الأشكال

الشكل 1: مجرى نهر الفرات في الجمهورية العربية السورية 43

الشكل 2: مبدأ عمل التنقيب عن البيانات 46

الشكل 3: التخصصات التي تلتقي في التنقيب عن البيانات 46

الشكل 4: إجرائية التنقيب في البيانات 47

الشكل 5: النموذج العام لمصفوفة الارتباك 55

الشكل 6: الصيغة النهائية لمجموعة البيانات بعد المعالجة 65

الشكل 7: المخطط التدفقي للتابع Main 67

الشكل 8: المخطط التدفقي للتابع About 68

الشكل 9: المخطط التدفقي للتابع About 69

الشكل 10: المخطط التدفقي للتابع Predict 70

الشكل 11: المخطط التدفقي للتابع result 71

الشكل 12: الواجهة الرئيسية للتطبيق 72

الشكل 13: واجهة معلومات التطبيق 1 73

الشكل 14: واجهة معلومات التطبيق 2 74

الشكل 15: واجهة تحليل النتائج 1 75

الشكل 16: واجهة تحليل النتائج 2 75

الشكل 17: واجهة تحليل النتائج 3 76
الشكل 18: واجهة تحليل النتائج 4 76
الشكل 19: واجهة التنبؤ قبل التنفيذ..... 77
الشكل 20: واجهة التنبؤ بعد التنفيذ..... 78
الشكل 21: وارد نهر الفرات بناءً على مدخلات المثال الأول 79
الشكل 22: وارد نهر الفرات بناءً على مدخلات المثال الثاني 80
الشكل 23: وارد نهر الفرات بناءً على مدخلات المثال الثالث..... 81
الشكل 24:وارد نهر الفرات بناءً على مدخلات المثال الرابع..... 82
الشكل 25: فئات مدخل "سعر برميل النفط" بالنسبة لعينات مجموعة البيانات (وارد نهر الفرات) 83
الشكل 26: فئات مدخل "الهطول المطري" بالنسبة لعينات مجموعة البيانات (وارد نهر الفرات). 84
الشكل 27: فئات مدخل "الموسم في السنة" بالنسبة لعينات مجموعة البيانات (وارد نهر الفرات) 85
الشكل 28: فئات مدخل "درجة الحرارة" بالنسبة لعينات مجموعة البيانات (وارد نهر الفرات) 85
الشكل 29: فئات وارد نهر الفرات بالنسبة لعدد العينات في مجموعة البيانات..... 86

فهرس الجداول

الجدول 1: مجموعة بيانات لمبيعات أجهزة الكترونية 56

الجدول 2: احتمالات شراء جهاز حاسب بحسب سمات الزبون المختلفة 58

الجدول 3: تحويل القيم الرقمي المتعلقة بسعر برميل النفط إلى فئات 62

الجدول 4: تحويل القيم الرقمي المتعلقة بدرجات الحرارة إلى فئات 62

الجدول 5: تحويل القيم الرقمي المتعلقة بالهطول المطري إلى فئات 63

الجدول 6: تحويل القيم الرقمي المتعلقة بوارد نهر الفرات إلى فئات..... 63

الجدول 7: مجموعة البيانات بعد المعالجة 64

الجدول 8: المكتبات المستخدمة في بناء تطبيق الويب الخاص بالمشروع..... 66

الجدول 9: مدخلات المثال الأول 78

الجدول 10: مدخلات المثال الثاني 79

الجدول 11: مدخلات المثال الثالث..... 80

الجدول 12: مدخلات المثال الرابع 81

الجدول13 : تدريب مجموعة البيانات باستخدام خوارزمية OneR بالطريقة الأولى..... 88

الجدول14 : تدريب مجموعة البيانات باستخدام خوارزمية OneR بالطريقة الثانية 89

الجدول15 : تدريب مجموعة البيانات باستخدام خوارزمية OneR بالطريقة الثالثة 90

الجدول16 : تدريب مجموعة البيانات باستخدام خوارزمية RandomTree بالطريقة الأولى 91

الجدول17 : تدريب مجموعة البيانات باستخدام خوارزمية RandomTree بالطريقة الثانية 92

الجدول18 : تدريب مجموعة البيانات باستخدام خوارزمية RandomTree بالطريقة الثالثة 93

الجدول19 : تدريب مجموعة البيانات باستخدام خوارزمية Naive Bayes بالطريقة الأولى 94

الجدول20 : تدريب مجموعة البيانات باستخدام خوارزمية Naive Bayes بالطريقة الثانية 95

الجدول21 : تدريب مجموعة البيانات باستخدام خوارزمية Naive Bayes بالطريقة الثالثة.....96

الجدول 22: خلاصة عمليات التدريب باستخدام الخوارزميات المختارة بثلاثة طرق مختلفة 97

المصطلحات

مصطلحات فنية

- **الإدارة المتكاملة للموارد المائية Integrated water resources management**

وهي عملية تشجع على التنسيق والإدارة للمياه والأراضي والموارد ذات الصلة، بغية تحقيق أقصى قدر ممكن من الرفاه الاقتصادي والاجتماعي الناجم على نحو عادل دون المساس باستدامة النظم الإيكولوجية الحيوية.

تستند الإدارة المتكاملة للموارد المائية إلى ثلاثة مبادئ تعمل معا كإطار عام :

- المساواة الاجتماعية: ضمان عدالة الوصول لجميع المستخدمين إلى كمية كافية من المياه ونوعيتها اللازمة للحفاظ على رفاه الإنسان .
- الكفاءة الاقتصادية: تحقيق أكبر قدر من الفائدة لأكبر عدد ممكن من المستخدمين من الموارد المالية والمائية المتاحة .
- الاستدامة الإيكولوجية: تتطلب الاعتراف بالنظم الإيكولوجية المائية كمستخدمين وأن يتم تخصيص ما يكفي للحفاظ على أدائها الطبيعي .

- **استدامة الموارد المائية Sustainability of water resources**

تُعرَّف الاستدامة بأنها ممارسة حماية الموارد الطبيعية لاستخدامها في المستقبل دون الإضرار بالطبيعة، تشمل التنمية المستدامة القضايا البيئية والاجتماعية والسياسية والاقتصادية التي تواجه الإنسان في

الوجود، والماء هو المورد الأكثر حيوية للعيش على هذه الأرض، يتم استغلال الموارد الطبيعية مع زيادة عدد سكان العالم وقد يهدد نقص هذه الموارد البشرية في المستقبل.

استدامة المياه هي جزء من الاستدامة البيئية. تتزايد أزمة المياه بشكل تدريجي في العديد من أماكن العالم بسبب استخدامات زراعية وصناعية والتوسع الحضري السريع، توفر أدوات وتقنيات استخراج البيانات منهجية قوية لفهم قضايا استدامة المياه باستخدام بيانات بيئية غنية وتساعد أيضًا في بناء نماذج للتحسين وإعادة الهندسة الممكنة. يوجد أعمال بحثية هامة يتم فيها تقديم مراجعة حول استخدام خوارزميات التعلم الخاضعة للإشراف أو غير الخاضعة للإشراف في قضايا استدامة المياه مثل تقييم جودة المياه ونظام جمع مياه الصرف واستهلاك المياه. ساعدت التقنيات المتقدمة أيضًا في حل مشكلات استدامة المياه الرئيسية.

الفشل في تحقيق الاستدامة المائية يعني صراعات أكبر في المستقبل لذا علينا إدارة المياه بشكل أكثر فاعلية.

- **السياسة المائية water policy**

هي الإطار الذي يتم من خلاله إدارة الموارد المائية واستنباط مجموعة القواعد التي تنظم ذلك، وتتضمن إصدار توجيهات فيما يخص الماء الصالح للاستخدام البشري وتأمينه وتوفير مصدر مستدام منه بحيث يتم ضمان استمرار الحياة والتطور البشري. بحيث تحقق المبادئ العامة كمرتكزات للسياسات المائية:

- مبدأ الاستدامة: ويقصد بهذا المبدأ سعي الدول والشعوب إلى تحقيق التنمية المستدامة ضماناً لحقوق الأجيال القادمة واستمرار الحياة على كوكب الأرض وبخاصة إن مورد المياه لا بديل عنه .
- مبدأ العدالة في التوزيع: يقصد به في هذا المجال تحقيق العدالة في مجال توفير المياه وتوزيعها كحصص على مستوى الأفراد والشرائح الاجتماعية وخصوصاً الضعيفة منها والقطاعات وتحقيق العدالة في التوزيع يجب ألا يقتصر على إمدادات مياه الشرب وإنما في كافة القطاعات .
- سلامة البيئة: بالحفاظ على الموارد الجوفية المتجددة وغير المتجددة من التلوث.

ويجب أن تهدف السياسات المائية العربية إلى تحقيق ما يلي:

- المحافظة على الحقوق السيادية الوطنية لموارد المياه القومية والمياه المشتركة مع الدول الأخرى وطبقاً لقواعد القانون الدولي والاتفاقيات المبرمة.
- تطوير السياسات والتشريعات المائية .
- الاستغلال الأمثل للموارد المائية لتحقيق التنمية المستدامة وفق مبدأ الميزة النسبية بهدف تحقيق الأمن المائي والغذائي على الصعيدين المحلي والاقليمي .
- ترشيد استثمار مصادر المياه الجوفية والاستغلال الأمثل لها في الزراعة والحفاظ عليها من التلوث والاستنزاف.

- **الموارد المائية Water Resources**

وتنقسم إلى قسمين هما:

المياه السطحية (كافّة أشكال المسطحات المائية فوق سطح الأرض؛ كالجداول، والأنهار، والبحيرات، والأراضي الرطبة، والمحيطات) والمياه الجوفية (المياه التي تتسرّب خلال طبقات الأرض من الحصى، أو الرمل، أو الصخور المكسّرة)

- **نوعية المياه water quality**

هي الخصائص الكيميائية والفيزيائية والبيولوجية للمياه المطلوبة لغرض معيّن أو هي مدى ملاءمة المياه للاستخدامات أو العمليات المختلفة، مثلاً القيود المفروضة على تركيزات المواد السامة لاستخدام مياه الشرب، أو القيود المفروضة على درجات حرارة المياه وحموضتها، ويستخدم هذا المصطلح بكثرة للدلالة على معايير لتقييم المياه

مصطلحات تقنية

- **البيانات الضخمة (Big Data):** يشير إلى البيانات الكبيرة جداً أو السريعة أو المعقدة بحيث يصعب أو يستحيل معالجتها باستخدام الطرق التقليدية، بمعنى آخر هي مجموعة من البيانات الضخمة في الحجم، التي تتولد بسرعة غير ثابتة وتزداد وتنمو بشكل كبير بمرور الوقت، ويصعب التعامل معها باستخدام أجهزة معالجة وتخزين البيانات التقليدية، ويرجع السبب في ذلك إلى أن مصدر البيانات ينتج كمية هائلة منها بشكل مستمر.

- **التصنيف (Classification):** شكل من أشكال تحليل البيانات التي تعمل على استخراج نماذج وأنماط تصف تلك البيانات وتصنفها، تقوم هذه النماذج التي تسمى مصنفات classifiers بتوقع الصنف الذي تنتمي له البيانات ويتسم الصنف بكونه فئوياً محدداً وغير مرتب.

- **التعلم الآلي (Machine learning):** هو فرع من فروع الذكاء الاصطناعي وأحد علوم الحاسوب، يركز بشكل كبير على استخدام البيانات (بيانات التدريب) والخوارزميات، من أجل محاكاة الطريقة التي يتعلم بها الإنسان والعمل على تحسينها بشكل تدريجي، لذلك فإن التعلم الآلي هو برنامج يظهر قدرة معرفية تتشابه إلى حد كبير مع قدرة الإنسان، ويهدف إلى جعل

أجهزة الكمبيوتر تعالج وتحلل مختلف المشاكل التي تواجهها بالطريقة التي يقوم بها الإنسان الطبيعي، من خلال الاستفادة من النماذج الإحصائية وخوارزميات التعلم. والهدف من التعليم الآلي جعل الأنظمة قادرة على التعلم والتطور بشكل تلقائي، من خلال التجربة، ودون الحاجة إلى القيام بعمليات برمجية.

- **التنبؤ (Prediction):** هو عملية بناء التوقّعات المدروسة لما سيحدث في المستقبل من قِبَل خبراء العلوم والرياضيّات، وذلك وفق عدّة معايير ومعلومات أوّليّة. والتنبؤ يكون في مُختلف العلوم والأنشطة كالاقتصاد، علم الاجتماع، علم الأحياء وغيرها من العلوم الأخرى.

- **التنقيب عن البيانات (Data Mining):** هو تنقيب عن البيانات لاكتشاف المعرفة، وهو عملية متطورة تقوم باستخلاص المعلومات المطلوبة والفعَّالة المخبأة، من كم هائل من البيانات طبقاً لأهداف مسبقة، وهي تكنولوجيا حديثة فرضت نفسها في عصر المعلوماتية، وهي القدرة على الاستكشاف والتركيز على أهم المعلومات في كتل البيانات الكبيرة وتحويلها من بيانات متراكمة وغير مفهومة إلى معلومات قيمة يمكن استغلالها والاستفادة منها لتصبح معرفة. كما تركز تقنيات التنقيب على الاستشعار وبناء التنبؤات المستقبلية واستكشاف الأنماط والارتباطات

والسلوك والاتجاهات مما يسمح بتحديد القرارات الصحيحة واتخاذها في الوقت المناسب ووضع الحلول المناسبة للمشكلات والتخطيط والتطوير والتحديث في جميع المجالات، في حين يضع البعض التنقيب عن البيانات كخطوات أساسية في عملية اكتشاف المعرفة.

- **التنقيب عن المعلومات المائية (Water Data Mining):** تحليل البيانات وتحويلها إلى معلومات للحصول على تصورات وأفكار جديدة تفيد في اتخاذ القرارات الصحيحة لتحقيق التنمية المستدامة للمياه وتساعد في المعالجة المناسبة للخلل الممكن في إدارة المصادر المائية.

- **الذكاء الاصطناعي (Artificial intelligence):** يشير (AI) إلى الأنظمة أو الأجهزة التي تحاكي الذكاء البشري لأداء المهام، والتي يمكنها أن تحسن من نفسها استنادًا إلى المعلومات التي تجمعها. ويتعلق بالقدرة على التفكير الفائق وتحليل البيانات أكثر من تعلقه بشكل معين أو وظيفة معينة، يهدف إلى تعزيز القدرات والمساهمات البشرية بشكل كبير بغض النظر عما يقدم صورًا عن الروبوتات العالية الأداء الشبيهة بالإنسان التي تسيطر على العالم.

- **الشبكات العصبية (Neural Networks):** تشبه في تركيبتها تركيبة عقل الانسان، فهي تعمل بنفس الطريقة في نقل ومعالجة وتحليل المعلومات وصولاً إلى الاستنتاجات واكتشاف الأنماط والتنبؤات، كما تهدف خوارزميات هذه الشبكات إلى محاكاة الذكاء البشري وإكساب الآلة بعض قدرات مشابهة للعقل الطبيعي.

- **شجرة القرار (Decision Tree):** هي نموذج استكشافي يظهر على شكل شجرة، يمثل كل فرع من فروعها سؤالاً تصنيفياً، وتهدف تقنية شجرة القرار إلى تقسيم قواعد البيانات بهدف معين سبق وأن تم تحديده. ويصبح وجود عنصر معين في إحدى المجموعات (الفروع)، هو نتيجة تحقيقه سلسلة الشروط الموضوعة وصولاً لهذا الفرع.

- **العنقدة (clustering):** عبارة عن تقسيم البيانات ضمن تجمعات أو مجموعات جزئية تكون النقاط ضمن التجمع الواحد أقرب ما يمكن لبعضها (من وجهة نظر أو خاصية ما)، وأكثر بعداً عن نقاط باقي التجمعات.

- **قواعد الاقتران أو الارتباط (Association Rules):** هي اكتشاف العلاقات المثيرة للاهتمام بين المتغيرات في قواعد البيانات الكبيرة. الغرض منها هو تحديد القواعد القوية المكتشفة في قواعد البيانات باستخدام بعض المقاييس. ويولد ذلك قواعد جديدة بتحليل المزيد من البيانات. الهدف النهائي، هو مساعدة الآلة في محاكاة وفهم قدرات الارتباط التجريدي من البيانات الجديدة غير المصنفة بافتراض وجود مجموعة بيانات كبيرة.

- **قواعد البيانات (Database):** هي مجموعة منظمة من البيانات المخزنة على نظام الكمبيوتر، وتدعم كمية كبيرة من البيانات، والتجهيز المتزامن، والعمليات الفعالة، وتستخدم على نطاق واسع، وهي غالبا ما تخضع للتحديثات.

- **مستودعات البيانات (Data Warehouse):** هو نوع خاص من قواعد البيانات، لكنه الأمثل للاستعلام والتحليل. وبما أن مستودع البيانات يستخرج البيانات من مصادر وتقارير مختلفة، فإنه يُمَّكن التوصل قرارات عن طريق التحليل. تستخدم لتخزين كميات ضخمة من البيانات لفترات طويلة، وهذه البيانات يمكن تحديثها ولكن لا يمكن تعديلها لأنها تستخدم فقط للتحليل والدراسة.

- **النظم الخبيرة (Expert systems):** هي برامج تُحاكي أداء الخبير البشري في مجال خبرة معين، وذلك عن طريق تجميع واستخدام معلومات وخبرة خبير أو أكثر في مجال معين. وجدت من أجل استخلاص خبرات الخبراء –خصوصاً في التخصصات النادرة – وضمها في نظام خبير يحل محل الإنسان ويساعد في نقل هذه الخبرات لأناس آخرين بالإضافة إلى قدرته على حل المشكلات بطريقة أسرع من الخبير البشري. وتتألف النظم الخبيرة من شقين: الأول يتم فيه بناء قاعدة المعرفة والثاني يتم فيه طرح الأسئلة من قبل المستخدم والحصول على النتائج من النظام الخبير.

- **البيانات (Data):** هي معلومات في صورة خام وهي الحقائق الأولية بدون أي تعديل أو وضع في إطار معين يجعلها ذات أهمية وقيمة، لأن البيانات تعتبر معرفة او معلومة ولكن في وضعها الخام بدون اي تعديل او إضافة، وتكون البيانات عادة في صورة ارقام او تكون على شكل رموز وهي ليست بالضرورة مرتبطة ببعض.

- **المعلومات (Information):** هي المخرجات التي تم إنتاجها من أي نظام بعد ادخال البيانات عليه لذلك يمكن تعريفها على أنها كمية من البيانات التي تم تنظيمها وتنسيقها في إطار يجعل البيانات

ذات قيمة، تم صياغتها لتعطي معنى مناسب ذو فائدة، ولابد ان تكون المعلومات متجانسة تعبر عن أفكار ومفاهيم غير متناقضة.

– **المعرفة (Knowledge):** هي الحصيلة التي تنتج من البيانات والمعلومات، فهي عبارة عن مزيج مرن من الخبرة المؤطرة والقيِّمة والمعلومات السياقية ورؤية الخبراء التي توفر إطارًا لتقييم ودمج الخبرات والمعلومات الجديدة، يمكن الجمع بينها وبين الخبرة البشرية، فالمعرفة توجد داخل العقول ويحصل عليها الأفراد بعد مجهود في المعالجة والتحليل. يمكن ايجاد تعريف آخر للمعرفة على انها القيام بتصرف معين جاء خلاصة لإدخال البيانات ثم معالجتها للحصول على المعلومات ثم وضعها في إطار معرفي لتتكون المعارف والخبرات، فهي مجموعة المفاهيم والمعلومات الواقعية، بما في ذلك الأنماط المترابطة، وهي القدرة على فهم موضوع معين يتم الحصول عليه من خلال الخبرة والدراسة.

القسم النظري

الفصل الأول: الإطار العام للبحث

1.1 المقدمة

تعد المياه أحد الموارد أهمية وينظر إلى إدارتها على أنها حاسمة لاستمرارية العمل ونموه، وتحتل ندرة المياه المرتبة الأولى في المخاطر العالمية المحدقة بالمجتمعات على مدار الأعوام العشر التالية وذلك عندما يتجاوز الطلب المحلي على المتاح المائي في المنطقة، وتعتبر الأنهار الرئيسية من مصادر المياه الرئيسية في العالم [1].

ويوصى باستخدام التكنولوجيا في القطاع المائي لإيجاد حلول لإدارة مصادر المياه وللتزود بها، للعرض والطلب على المدى الطويل وذلك بالاعتماد على الأتمتة في تنفيذ المهام وتخزين البيانات في الحاسوب بدلاً من الطرق التقليدية.

وأحد الطرق التي يمكن بها نرفع مستوى إدارة مصادر المياه هي تحليل قواعد البيانات المتوفرة والتي تحتوي عدد لا بأس به من معلومات عن هذه المصادر وتدفقها وكمية الوارد فيها وغير ذلك. ويتم ذلك باستخدام أحد فروع التنقيب عن المعلومات وهو التنقيب عن الموارد المائية وتحديد توزيع المياه على المشاريع والمنشآت [2].

إن الهدف الأول من هذا البحث هو إظهار إمكانية الاستفادة من خوارزميات التنقيب عن المعلومات في القطاع المائي لتحليل واردات نهر الفرات للوصول إلى الإدارة المثلى لهذا المصدر المائي بالإضافة لبناء أداة عامة وسهلة الاستخدام (موقع انترنت Web App) لأتمتة عملية تحليل الواردات المخزنة مع إخفاء تفاصيل عملية التحليل عن المستخدم.

2.1 المشكلة العلمية ومبررات مشروع البحث

يوجد ثلاث عوامل رئيسية تؤثر على الخطط الموضوعة لاستثمار المورد المائي المشترك بين عدة دول:

1- ظروف طبيعية منها الظروف المناخية.

2- ظروف اقتصادية وسياسية عالمية تؤثر على ادرة المورد المائي المشترك.

3- سياسات تتعلق بالدول المشتركة لنهر الفرات، ومدى احترامها للالتزامات الدولية المبرمة معها (الالتزام بالاتفاقيات والبروتوكولات الموقعة بين الدول المتشاركة لاقتسام المورد المائي المشترك).

وترتبط العوامل الثلاثة السابقة مع بعضها البعض ارتباطاً وثيقاً بشكل عام وبالنسبة لنهر الفرات بشكل خاص حيث أنه نتيجة ما ذكر أعلاه يتم تمرير كمية من تدفقات نهر الفرات قد تقل أو تكثر عما هو متفق عليه وبالتالي يتعذر نتيجة هذا التذبذب في تدفقات النهر من وضع خطة مناسبة لإدارة مياه النهر بالشكل الأمثل وبالتالي لا يمكن وضع خطط زراعية وتنموية تقوم على ترشيد استخدام الموارد المائية.

3.1 الهدف من البحث

استخدام لغة الآلة في تحديد الوارد المائي المتوقع لنهر الفرات إلى أراضي الجمهورية العربية السورية عن طريق نظام تنبؤ وبناء أداة لأتمتة تحليل هذا الوارد المتوقع وعلاقته مع عوامل عدة: الهطولات المطرية والثلجية – درجات الحرارة – سعر النفط العالمي – الموسم.

4.1 أهمية البحث

ابراز دور التقنيات والأساليب الحديثة وإمكانية تطبيقها في إدارة وتخطيط الموارد المائية بشكل عام وفي منطقة الدراسة بشكل خاص.

5.1 النتائج التطبيقية المتوقعة من البحث والجهات المستفيدة منه

بناء نظام للكشف والتنبؤ عن قيم الواردات المائية لنهر الفرات إلى الجمهورية العربية السورية سيقدم قيمة مضافة لإدارة الموارد المائية تنعكس ايجاباً من الناحية الاجتماعية والاقتصادية ويرفع الكفاءة التخطيطية على مياهه مما يحقق الاستفادة القصوى في المشاريع المقامة عليه.

6.1 الجهات المستفيدة

- وزارة الموارد المائية بشكل خاص كونها المعنية بإدارة الموارد المائية في الجمهورية العربية السورية.
- وزارة الزراعة من خلال تحسين إدارة التخطيط الزراعي وزيادة الانتاج.
- وزارة الإسكان والمرافق من خلال تحسين التخطيط الإقليمي والعمراني والصناعي.

7.1 الدراسات السابقة

1.7.1 الدراسة الأولى

بعنوان **توقع جودة المياه الجوفية باستخدام خوارزميات التعلم الآلي في R** [3]، تتناول الدراسة الخصائص الفيزيائية والكيميائية لنوعية المياه الجوفية في مدن Ranipet, Arcot, Walljah pet في الهند، حيث تم جمع عينات المياه من مختلف الآبار الجوفية، لغرض دراسة جودة المياه (تستخدم الآبار التي تم جمع العينات منها على نطاق واسع لأغراض الشرب)، ومن ثم تم تقدير معايير نوعية المياه مثل PH ، TDS، EC، الكلوريد ، الكبريتات ، النترات ، الكربونات ، البيكربونات، أيونات المعادن ، العناصر النزرة.

ركز البحث على التنبؤ بنوعية المياه باستخدام خوارزمية مصنف التعلم الآلي C5.0 و Naïve Bayes و Random Forest مع أداة تحليل البيانات R لإنشاء نموذج تنبؤي فعال لتحليل نوعية المياه الجوفية يتنبأ بما إذا كانت المياه "عالية" أو "منخفضة" الجودة لأغراض الشرب بناءً على معايير جودة المياه، أنتج Naïve Bayes و Random Forest نتيجة أفضل بالدقة وخطأ اقل في التصنيف.

وقد كان المقترح استخدام خوارزمية التصنيف في المستقبل مع مجموعة بيانات ممتدة لتحليل نوعية المياه الجوفية، ومن ثم المعالجة المناسبة للمياه ضرورية لحماية صحة المجتمع.

2.7.1 الدراسة الثانية

بعنوان **تطبيق المصنف على أساس تقنيات التنقيب في البيانات في تشغيل إمدادات المياه** [4]، في هذه الدراسة قام الباحثون بتطبيق تقنية التنقيب عن البيانات لاستخراج قواعد تشغيل إمدادات المياه، وتم اختيار خمس سمات مميزة – كمية التخزين، ورقم فترة التشغيل ، والطلب على المياه ، والجريان السطحي ، والسنة الهيدرولوجية – كمجموعة بيانات ، وتم تطبيق هذه السمات المميزة لبناء علاقة تعيين مع وضع التشغيل الأمثل المحسوب بواسطة البرمجة الديناميكية (DP) وتم اختيار شبكة عصبية Levenberg–Marquardt (LM) وشجرة التصنيف والانحدار (CART) كخوارزميات استخراج البيانات لبناء مصنف الشبكة العصبية LM ومصنف شجرة القرار CART ، على التوالي، من أجل التحقق من تأثير التصنيف لـ LM و CART ، تم تطبيق المصنفين على التعرف على وضع التشغيل لخزان Heiquan ، الذي يقع في مقاطعة Qinghai في الصين. بلغت دقة المصنفين 73.6٪ و 86.9٪ لعينة التدريب ودقتها 65.8٪ و 83.3٪ على التوالي لعينة الاختبار وهو ما يشير إلى أن نتيجة التصنيف لمصنف CART أفضل من تلك الخاصة بمصنف الشبكة العصبية .LM ، وبالتالي تم اختيار مصنف CART لتوجيه عملية إمداد المياه ذات السلسلة الطويلة مقارنة بنتيجة العملية مع مخطط التشغيل الآخر ، تُظهر النتيجة أن مؤشر عجز المياه في CART هو الأقرب في الغالب إلى مخطط DP ، مما يشير إلى أن مصنف CART يمكنه توجيه عملية الإمداد بالمياه بشكل فعال.

3.7.1 الدراسة الثالثة

بعنوان **التنبؤ بجودة المياه باستخدام خوارزميات الذكاء الاصطناعي** [5]، في هذا العمل، تم تطوير خوارزميات الذكاء الاصطناعي (AI) المتقدمة للتنبؤ بمؤشر جودة المياه (WQI) ، وتصنيف جودة المياه (WQC) .

للتنبؤ بمؤشر جودة المياه WQI، تم استخدام نماذج الشبكات العصبية الاصطناعية، وهي الشبكة العصبية غير الخطية الانحدار الذاتي (NARNET) وخوارزمية التعلم العميق للذاكرة طويلة المدى (LSTM).

بالنسب للتنبؤ بتصنيف جودة WQC تم استخدام ثلاث خوارزميات للتعلم الآلي، وهي آلة متجه الدعم (SVM) ، وأقرب جوار (KNN) ، وNaive Bayes .

تحتوي مجموعة البيانات المستخدمة على 7 معلمات مهمة، وتم تقييم النماذج المطورة بناءً على بعض المعلمات الإحصائية. أظهرت النتائج أن النماذج المقترحة يمكن أن تتنبأ بدقة WQI وتصنف جودة المياه وفقًا للمتانة الفائقة. كما أظهرت نتائج التنبؤ أن نموذج NARNET كان أداؤه أفضل قليلاً من LSTM للتنبؤ بقيم WQI وأن خوارزمية SVM قد حققت أعلى دقة (97.01٪) لتنبؤ WQC .علاوة على ذلك، حقق طرازا NARNET و LSTM دقة مماثلة لمرحلة الاختبار مع اختلاف طفيف في معامل الانحدار:

$$(R_{NARNET} = 96.17\% \text{ and } R_{LSTM} = 94.21\%)$$

بعد فحص متانة وكفاءة النموذج المقترح للتنبؤ بمؤشر جودة المياه، في العمل المستقبلي، يمكن تنفيذ النماذج المطورة للتنبؤ بجودة المياه لأنواع مختلفة من المياه حيث يمكن أن يساهم هذا النوع من الأبحاث الواعدة بشكل كبير في إدارة المياه.

4.7.1 الدراسة الرابعة

بعنوان **نموذج تصنيف لتحليل جودة المياه باستخدام شجرة القرار** [6]، تقدم هذه الدراسة نموذج تصنيف باستخدام شجرة القرار لغرض تحليل بيانات جودة المياه من مقاطعات مختلفة في كينيا، لضمان حصول المواطنين على مياه نظيفة، يمكن أن يؤدي تطبيق شجرة القرار كطريقة لاستخراج البيانات للتنبؤ بالمياه النظيفة بناءً على معايير جودة المياه إلى تسهيل عمل فني المختبر من خلال توقع عينات المياه التي يجب أن تنتقل إلى الخطوة التالية من التحليل، تم استخدام البيانات الثانوية من معهد المياه في كينيا لإنشاء النموذج، تم تنفيذ نموذج البيانات في برنامج .WEKA ثم تطبيق التصنيف باستخدام شجرة القرار لتصنيف / توقع المياه النظيفة وغير النظيفة. لعب تحليل قلوية المياه ومستوى الأس الهيدروجيني والتوصيل دورًا رئيسيًا في تقييم جودة المياه، تم استخدام خمسة مصنفات لشجرة القرار وهي J48 و LMT و Random Forest و Hoeffding Tree و Decision Stump لبناء النموذج ومقارنة الدقة. حصلت شجرة القرار J48 على أعلى دقة بنسبة 94٪ مع دقة أقل بنسبة 83٪ فيDecision Stump .

يمكن استخدام مصنف شجرة القرار المزود بمعايير جودة المياه التي حددتها منظمة الصحة العالمية و KEBS للتنبؤ فيما إذا كانت مياه الشرب صالحة أم لا.

5.7.1 الدراسة الخامسة

بعنوان **تطبيقات طرق التعلم الآلي المختلفة للتنبؤ بمستوى المياه** [7]، تتميز آيسلاندا بالتفرد بأن معظم الكهرباء المنتجة في أيسلندا تقريبًا يمكن أن تُعزى إلى مصادر الطاقة المتجددة، والطاقة المائية هي أحد هذه المصادر، إن تدفق المياه في محطات الطاقة في أيسلندا غير متوقع ويختلف بشكل كبير بين المواسم. مع وجود نموذج تنبؤ جيد يتنبأ بمستوى المياه في الخزانات، سيكون من الممكن استخدام المياه بشكل أفضل عن طريق تقليل الفائض.

في هذه الورقة، تم استخدام أربع طرق مختلفة للتنبؤ بمستوى المياه: multiple linear regressionالانحدار الخطي المتعدد– random forest الغابات العشوائية– support vector regression انحدار ناقلات الدعم– artificial neural networks الشبكات العصبية الاصطناعية.

تستند جميع نماذج التنبؤات إلى بيانات تاريخية من مكتب الأرصاد الجوية الأيسلندي واحتوت البيانات على قياسات التدفق ومستويات المياه ومتغيرات الطقس المختلفة من 2009 إلى 2017. كما تم تقسيم البيانات إلى مجموعة تدريب تستخدم لتعيين معلمات النماذج ومجموعة اختبار تستخدم لقياس أداء النماذج، كان الاختلاف الرئيسي بين النماذج هو التوقع لأعلى وأدنى مستويات المياه، وقد كان النموذج الذي أسفر عن أدنى خطأ هو نموذج random forest المستند إلى البيانات اليومية.

6.7.1 الدراسة السادسة

بعنوان **تحليل أدوات التنقيب عن البيانات المستخدمة في إدارة الموارد المائية في نهر دجلة** [8]، قام الباحثون بدراسة مقارنة بين عدد من أدوات التنقيب عن البيانات وأدوات اكتشاف المعرفة وحزم البرمجيات التي تُستخدم في أنظمة دعم القرار (DSS) لإدارة الموارد المائية. وتناولت الدراسة تقييم جودة مياه نهر دجلة في منطقة مدينة بغداد.حيث تم تحديد إجراءات بديلة لتحسين استدامة نوعية المياه .

تستخدم طريقة جودة المياه مجموعة من ثلاثة عشر معيارًا لجودة المياه وفقًا لتحليل مؤشر جودة المياه (WQI) باستخدام أداة التنقيب (WEKA). وقد أظهرت النتائج أن مياه النهر ليست ضمن معايير الجودة القياسية لمعايير منظمة الصحة العالمية والمعايير العراقية.

7.7.1 الدراسة السابعة

بعنوان **كيف يمكن للبيانات الضخمة والتعلم الآلي أن تفيد إدارة البيئة والمياه** [9]، الغرض من هذا البحث هو فحص إمكانات وفوائد البحث القائم على البيانات في **إدارة البيئة والمياه EWM** ، وتقديم ملخص للمفاهيم والأساليب الرئيسية في البيانات الضخمة وتعلم الآلة، وتقديم مراجعة منهجية للتطبيقات الحالية، ومناقشة القضايا والتحديات الرئيسية، والتوصية باتجاهات البحث المستقبلية.

حيث تتمتع تقنيات البيانات الضخمة Big Data والتعلم الآلي (ML) بالقدرة على التأثير على العديد من جوانب إدارة البيئة والمياه (EWM) مثل التنبؤ بالطقس، وإدارة الكوارث، وأنظمة إدارة المياه والطاقة الذكية، وحيث أن الاستشعار عن بعد وفرت البيانات الضخمة بدقة عالية، ولكنها تتطلب أيضًا أشكالًا

جديدة من معالجة المعلومات وتخزينها واسترجاعها بالإضافة إلى اجراء التحليلات عليها. يشير MLالتعلم الآلي على نطاق واسع إلى خوارزميات الكمبيوتر التي يمكنها التعلم تلقائيًا من البيانات. وتفيد التطورات التكنولوجية بشكل كبير في EWM ، خاصة لأن العديد من تطبيقات EWM مثل الإنذار المبكر بالفيضانات تتطلب القدرة على استخراج معلومات مفيدة من كمية كبيرة من البيانات بطريقة مستقلة وفي الوقت الفعلي.

8.7.1 الدراسة الثامنة

بعنوان **التنبؤ بالفيضانات باستخدام التعلم الآلي** [10]، تقدم هذه الدراسة أكثر طرق التنبؤ طويل الاجل وقصير الأجل للفيضانات. علاوة على ذلك، تم التحقيق في تحسين جودة نماذج التنبؤ بالفيضانات. من بينها: تحليل البيانات، ومجموعة الخوارزمية، وتحسين النموذج، وهي الاستراتيجية الأكثر فاعلية في تحسين طرق التعلم الآلي، يمكن استخدام هذا المسح كدليل إرشادي لعلماء الهيدرولوجيا وعلماء المناخ لمساعدتهم في اختيار الطريقة المناسبة وفقًا لاستنتاجات مهمة للتنبؤ.

تعتبر الفيضانات من بين أكثر الكوارث الطبيعية تدميراً، وهي معقدة للغاية من حيث النماذج. ساهم البحث عن تطوير نماذج التنبؤ بالفيضانات في الحد من المخاطر، واقتراح السياسات، وتقليل الخسائر في الأرواح البشرية وتقليل الأضرار التي تلحق بالممتلكات المرتبطة بالفيضانات.

ساهمت أساليب التعلم الآلي (ML) بشكل كبير في تطوير أنظمة التنبؤ التي توفر أداءً أفضل وحلولًا فعالة من حيث الكلفة.

يهدف الباحثون من خلال تقديم أساليب تعلم الآلة الجديدة إلى اكتشاف نماذج تنبؤ أكثر دقة وكفاءة، من خلال تحليل نوعي للقوة والدقة والفعالية والسرعة بشكل خاص لتقديم نظرة عامة شاملة على استخدام خوارزميات التعلم الآلي المختلفة في هذا المجال.

9.7.1 الدراسة التاسعة

بعنوان **تحليل مقارن لمصادر المياه الذكية باستخدام أدوات استخراج البيانات** [11]، تتوقع المدن نموًا سكانيًا هائلاً، وبالتالي فإنها ستحتاج إلى بنية تحتية جديدة وذكية لتلبية احتياجات مواطنيها وشركاتها. في هذا المشروع تمت صياغة نموذج تدفق المياه الجوفية لمدينة الدراسة باستخدام بيانات الإدخال ، مثل موقع مصدر مائي وظروف الحدود المناسبة، يحتاج هذا المشروع إلى جمع البيانات من مصادر مختلفة وتحليل تلك البيانات باستخدام بعض أدوات جمع البيانات للتنبؤ أو عملية صنع القرار وبعد جمع البيانات المختلفة ، كان المطلوب الحفاظ عليها وتطبيق التحويل والمعالجة المسبقة لمجموعات البيانات الكبيرة لأدوات استخراج البيانات مثل WEKA و Orange و R Studio و Tinn R و R للدراسة والمقارنة حول تحليل المصادر المائية، مع إمكانية التنقيب عن البيانات المتكاملة ونماذج البحث الأولية للطرق المطورة حديثًا، وتمت مناقشة النتائج حول مختلف أدوات التنقيب عن البيانات المتاحة ومقارنة أدواتها.

10.7.1 الدراسة العاشرة

بعنوان **نموذج لكشف الاحتيال في استهلاك المياه في مدينة غزة بالاعتماد على تقنية تنقيب البيانات ونظام الفاتورة المحوسبة** [12]، يقوم البحث بالمساعدة في الكشف الذين يمارسون الاحتيال في سحب كميات المياه عن طريق ايجاد أدوات للقياس للكشف عن عمليات الاحتيال في الاستهلاك المرتبط بعداد الكتروني أو ميكانيكي لحساب كمية الاستهلاك ، عن طريق تطبيق تقنية تنقيب البيانات على البيانات الناتجة عن الأنظمة، كما يهدف إلى ايجاد طريقة مناسبة من طرق تنقيب البيانات المتخصصة للتعامل مع نظام الفوترة المالية الخاص بإدارة حسابات مشتركين المياه لمدينة غزة .

تم تصميم النموذج بحيث يكون قادراً على تحديد مجموعة المشتركين المشبوهة، وتم اخضاعه لإحدى عمليات التقييم وصولاً إلى مستوى أداء ودقة مقبولين، حيث كان من المتوقع أن يساعد فريق الكشف عن سرقات المياه حيث سيزيد معدل الكشف من 5% إلى 85%.

تضمن البحث طريقة من طرق تنقيب البيانات التي تشمل اختيار واستخراج الملامح أو العناصر ذات الصلة من البيانات التاريخية الخاصة باستهلاك المياه للمشتركين مع تطبيق المصنف SVM على الملفات الخاصة بأحمال أو استهلاكيات المشتركين.

11.7.1 الدراسة الحادية عشر

بعنوان **نظام الإنذار المبكر بالفيضانات الحضرية ومخاطر جودة المياه** [13]، قام الباحثون بتطبيق الشبكات العصبية الاصطناعية (ANNs) كنماذج تستند إلى البيانات (DDMs) للتنبؤ بالفيضانات الحضرية في الوقت الفعلي بناءً على رادار الطقس و/أو بيانات مقياس هطول الأمطار. يتم تكوين

ANNالمتأخر زمنيًا للتنبؤ بالفيضانات في عقد الصرف الصحي والمصارف بناءً على معلمات الإدخال بما في ذلك هطول الأمطار. في حالة عدم وجود بيانات الفيضانات المرصودة، يمكن استخدام محاكي هيدروديناميكي للتنبؤ بمستويات الفيضان في العقد ذات الأهمية في شبكات الصرف الصحي وبالتالي توفير البيانات المستهدفة لتدريب واختبار ANN .

يعمل النموذج، كبديل سريع لجهاز المحاكاة الهيدروديناميكي ويمكن بالتالي استخدامه كجزء من نظام الإنذار المبكر للفيضانات الحضرية (EWS) . يتم ادخال هطول الأمطار المتوقعة على أحواض مائية، لتمديد أوقات التنبؤ إلى مستويات مفيدة من الناحية التشغيلية. يتم تنفيذ كل من مخططات تصنيف مستوى الفيضان التماثلية وشدة الفيضانات.

باستخدام شبكات ANN للتنبؤ الآني لهطول الأمطار بناءً على العلاقة بين بيانات الرادار وتاريخ هطول الأمطار المسجل ؛ تم وصف مخطط استخراج الميزات. سيسمح هذا بتدرج شبكتي ANN للتنبؤ بالفيضانات في الوقت الفعلي بناءً على رادار الطقس. وتم استخدام نفس المنهجية لتوقع جودة مياه (WQ) ومخاطرها.

الفصل الثاني: المياه ونهر الفرات

1.2 مقدمة

عادة يتم الاستفادة من آراء الخبراء الفنيين والاستعانة بوسائل القياس والاستشعار عن بعد في جمع البيانات والمعطيات الدقيقة عن المواقع وتحديثها على نحو مستمر لوضعها تحت تصرف أصحاب القرار عند رسم السياسات المائية واعداد الخطط بعيدة المدى، وتشكل الإدارة المتكاملة للموارد المائية خياراً استراتيجياً يوصى به لكافة أنشطة التنمية الاجتماعية والاقتصادية المتصلة بالمياه.

غير أن تزايد الطلب على المياه في ظل موارد محدودة وأحيانا غير متجددة وظهور أنماط حياتية وصناعية جديدة أدى إلى تصاعد كبير في الاستهلاك المائي في الوقت الذي تتزايد فيه الكثافة السكانية والتصحر وتلوث البيئة وتأثير التغيرات المناخية من فيضانات وجفاف، ويزداد الأمر تعقيداً عندما يتعلق بالبلدان التي تتشارك نفس المورد أو المجرى المائي، ونتيجة لهذه العوامل فقد طرأت على الموارد المائية تغيرات كمية ونوعية انعكست سلباً على تأمين الامداد بالمياه أو على استدامتها [14].

بالإضافة لما سبق فإن نشوء الحدود السياسية للدول رافقه ظهور مشكلة تحديد حق الدول في اقتسام الأنهار المشتركة وقيام دولة المنبع بإنشاء مشاريع مائية وكهرومائية في أعالي الحوض للنهر دون مراعاة مصالح الشركاء أدنى الحوض.

2.2 الوضع المائي في سوريا

تعتبر الجمهورية العربية السورية بلداً محدود الموارد المائية وتصنف ضمن المناطق الجافة وشبه الجافة ويتراوح معدل الهاطل المطري فيها بحسب المناطق بين 100ملم–1500 ملم.

يزداد الطلب على المياه من قبل كافة القطاعات كانعكاس لتزايد عدد السكان وما يرافقه من متطلبات التنمية الاقتصادية والاجتماعية، وبالتالي يوجد عجز دائم بالموازنة المائية في سوريا حيث يتم تغطية هذا العجز من احتياطي المياه الجوفية مما أدى لهبوط مناسيبها وزيادة ارتفاعات وتكاليف الضخ وجفاف بعض الينابيع وانخفاض تصريف عدد كبير منها [15].

1.2.2 التحديات الرئيسية في قطاع المياه في الجمهورية العربية السورية

ان ضمان استدامة الموارد المائية ضمن المعايير الفنية والاقتصادية–الاجتماعية هو هدف مركزي ومسؤولية مشتركة لكافة القطاعات المعنية، ويواجه تحقيقه بعض الصعوبات والتحديات وأهمها:

1. التغيرات المناخية: تؤدي ظواهر التغيرات المناخية (ارتفاع في درجات الحرارة – انخفاض الهطولات المطرية–الجفاف وغير ذلك) إلى تراجع كمي ونوعي في الواردات المائية المتجددة المتاحة للاستخدام في مجمل أراضي الجمهورية العربية السورية ودول الجوار.
2. ازدياد الطلب على الموارد المائية: ارتفاع معدلات النمو السكاني مع ازدياد النشاط الاقتصادي المرافق يفرض ضغوطاً إضافية على الموارد المائية مع محدوديتها قياساً بحجم الطلب المتزايد عليها.

3. الموارد المائية المشتركة: حوالي 60% من الموارد المائية في الجمهورية العربية السورية تنبع خارج أراضيها، ونتيجة هشاشة الأوضاع السياسية في المنطقة عموماً وعدم الالتزام بالاتفاقيات والتفاهمات الدولية من قبل بعض الدول المتشاركة يؤثر سلباً مؤدياً لعدم استقرار الواردات نظراً لتحكم دول المنبع فيها وعدم قدرة دول المصب على التحكم أو التنبؤ بكمية ونوعية المياه الواردة.
4. التلوث وتدهور نوعية المياه: ويتمثل ذلك في تصريف مياه الصرف الصحي والصناعي للمجاري المائية والأودية وكذلك تأثير المبيدات والأسمدة المتسربة للمياه الجوفية مما يؤدي إلى تدهور المصادر المائية.
5. الاستخدام غير المستدام للموارد المائية: ويتجلى ذلك في التخطيط لمساحات مروية تفوق احتياجاتها المتاح المائي مما أدى إلى آثار سلبية وخاصة على المياه الجوفية.

مما سبق نجد أن الاستثمار في المشاريع المائية للأغراض الزراعية وغيرها هو استثمار عالي الكلفة [16]. وبالتالي فإن تقدير كميات الواردات المائية هو في منتهى الأهمية للمخطط المائي في سوريا تجنباً للهدر وضياع الفرص التنموية والاستثمارية المرافقة.

3.2 أهمية نهر الفرات

نهر الفرات هو أحد الأنهار الكبيرة في جنوب غرب آسيا (الشكل 1) وأكبر نهر في الصفيحة العربية، وينبع النهر من جبال طوروس في تركيا وتنضم إليه فروع عديدة قبل مروره في أراضي الجمهورية العربية السورية ليجري في الاراضي العراقية.

ويدخل نهر الفرات في أراضي الجمهورية العربية السورية عند مدينة جرابلس، ثم يمر في محافظة الرقة ويتجه بعدها إلى محافظة دير الزور، ويخرج منها عند مدينة البوكمال. ويبلغ طول نهر الفرات من منبعه في تركيا حتى مصبه في شط العرب في العراق حوالي 2940 كم منها 610 كم في سوريا . ولنهر الفرات والسدود المقامة عليه أهمية كبرى في منظومة مياه الري والشرب في سوريا إضافة إلى محطات توليد الطاقة الكهرمائية المقامة عليه [16].

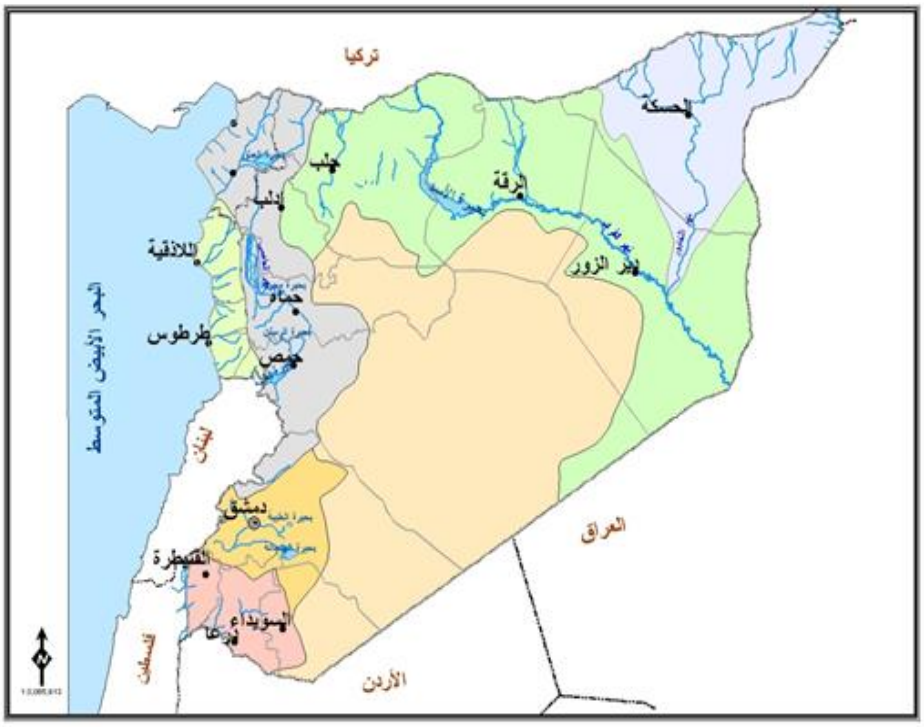
ويعتبر نهر الفرات نهراً دولياً ويتم بموجب اتفاقيات ثنائية أومؤقتة تقاسم مياه النهر بين ثلاثة دول وهي الجمهورية العربية السورية والعراق وتركيا. وتحدد كمية المياه الواردة إلى الجمهورية العربية السورية المؤقتة وفق ما يلي:

1. بروتوكول 1987 مع تركيا لتمرير 500 م3/ثا كحد أدنى من مياه النهر إلى الجمهورية العربية السورية في جرابلس، أي ما يعادل كمية 15.8 مليار م3 سنوياً.
2. اتفاقية عام 1989 مع العراق وتنص بأن تمرر الجمهورية العربية السورية للعراق 58% من المياه الواردة في جرابلس، وتعادل حصة الجمهورية العربية السورية المقدرة بـ 42% من واردات الفرات (أي 210 م3/ثا) حوالي 6.622 مليار م3 سنوياً.

- يتم استثمار الكميات الممررة زيادة عن 500 م3/ثا بإنتاج الكهرباء.
- إن ما يميز الطاقة المنتجة بمحطات الفرات هو رخص تكاليف إنتاجها ونظافتها البيئية وإنتاجها بساعات ذروة الطلب.
- بلغت مساهمة منظومة الفرات بإنتاج الطاقة الكهربائية في سوريا منذ بدء سد الفرات بإنتاج الطاقة حتى نهاية 2012 حوالي 16.5% من إجمالي الإنتاج.

يتطلب التخطيط لتنفيذ المشاريع على نهر الفرات إعداد موازنة مائية للنهر حسب الواردات إلى الجمهورية العربية السورية والمقدرة بـ 6.622 مليار م3، مع ضمان استمرار هذه التدفقات.

بالمقابل يتمتع الجانب التركي بقدرة تنظيمية عالية على مجرى النهر تعطيه التحكم الكامل بالواردات إلى الجمهورية العربية السورية والعراق، وقد يتصرف الجانب التركي بناء على أجندة سياسية أو تحت التأثيرات المناخية [17].



الشكل 1: مجرى نهر الفرات في الجمهورية العربية السورية

الفصل الثالث: الذكاء الصنعي

1.3 تطور البيانات

في الوقت الحاضر، تعتمد جميع التطبيقات قواعد البيانات Databases ومستودعات البيانات data warehouses لتخزين كمية كبيرة منها؛ ومع ازدياد حجم هذه البيانات وبمرور الوقت يتم اللجوء إلى استخدام المزيد من قواعد البيانات .

تحتوي البيانات عادة على الكثير من المعرفة المفيدة Useful knowledge التي يمكن استغلالها وتحليلها؛ يمكن أن تكون هذه المعرفة أي نوع من القواعد أو الانتظام أو الأنماط أو القيود، لذا فإن اكتشاف المعرفة ضروري لفهم البيانات واستخدامها.

كما أدى الانتشار الواسع لتقانات المعلومات وسهولة إتاحتها إلى تضخم حجم المعلومات بصورة لم يشهدها التاريخ من قبل، عند التطرق إلى البيانات الضخمة فإن ذلك يعني كميات لا يمكن تخيلها من البيانات المتعددة الأنواع والمصادر لحجم يتجاوز المئات من التيرابايت أو حتى البيتابايت.

فما أهمية هذه البيانات في ظل وجود حقيقة تشير إلى أن المعلومات المنظمة من هذه البيانات لا تشكل إلاّ جزءً ضئيلاً يصل إلى 10% مقارنة بالمعلومات غير المنظمة والتي تشكل الباقي، مما يؤدي إلى ازدياد الحاجة إلى تطوير أدوات تمتاز بالقوة لتحليل البيانات واستخراج المعلومات والمعارف منها، فالأساليب التقليدية والاحصائية لا تستطيع التعامل مع هذا الكم الهائل من البيانات لذلك نلجأ لاستخدام أدوات ذكية لمعالجتها [18].

من هنا ظهر ما يُسمى باستخراج البيانات أو التنقيب عن البيانات Data Mining كتقنية تهدف إلى استنتاج المعرفة من كميات هائلة من البيانات، تعتمد على الخوارزميات الرياضية والتي تعتبر أساس في التنقيب عن البيانات وهي مستمدة من العديد من العلوم مثل: علم الإحصاء والرياضيات والذكاء الاصطناعي والنظم الخبيرة وعلم التعرف على الأنماط وعلم الآلة وغيرها من العلوم التي تعتبر من العلوم الذكية وغير التقليدية.

يعتبر التنقيب عن البيانات Data Mining أو Knowledge Discovering DataBase أحد الحلول الناجحة لتحليل كميات ضخمة من البيانات وذلك بتحويلها من مجرد معلومات متراكمة وغير مفهومة (بيانات) إلى معلومات يمكن استغلالها والاستفادة منها وذلك باستخدام خوارزميات قابلة للتوسع والتكيف مع الكميات المتزايدة من البيانات في البحث عن أنماط معرفية ذات معنى . فهو علم يهتم باستخراج الأنماط والنماذج غير المعروفة مسبقاً أو المخبأة في قواعد البيانات الضخمة بهدف فهم أعمق للبيانات المتوفرة من أجل اتخاذ القرارالأمثل [19] .

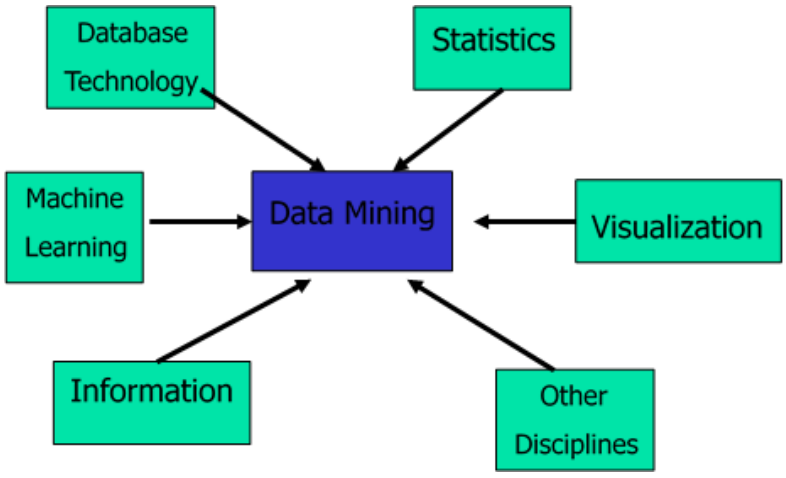
2.3 التنقيب البيانات

يُعرّف التنقيب في البيانات (الشكل 2) بأنه إجراء لاستخراج المعلومات من مجموعات ضخمة من البيانات. بعبارة أخرى، يمكننا القول أن التنقيب في البيانات هو استخراج معرفة ذات قيمة من البيانات.

الشكل 2: مبدأ عمل التنقيب عن البيانات

كما يُعرف التنقيب عن البيانات على النحو التالي :استخراج معلومات أو أنماط مثيرة للاهتمام (غير تافهة، ضمنية، غير معروفة سابقًا وربما مفيدة) من البيانات في قواعد البيانات الكبيرة، فهذا يعني أننا يجب أن نجد معرفة ذات قيمة في قاعدة بيانات ضخمة حيث تكون هذه المعرفة معقدة وليس من السهل العثور عليها من قبل البشر [20].

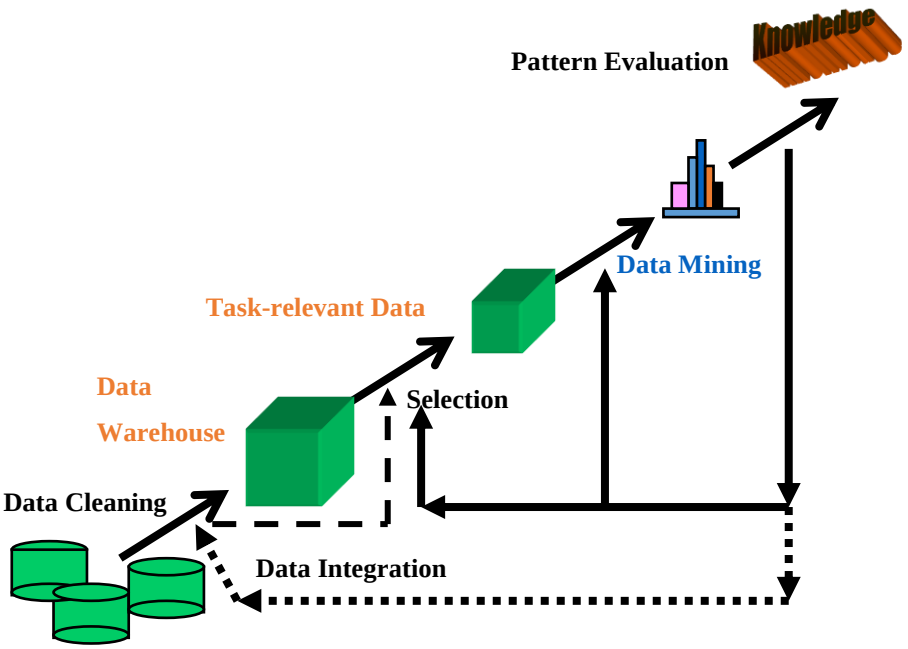
تأتي تقنيات استخراج البيانات من ثلاثة مجالات أكاديمية: الإحصاء، والتعلم الآلي، وتكنولوجيا قواعد البيانات، ويجب الأخذ بعين الاعتبار أن هناك أنماط جديدة وغير معروفة سابقاً قد تبرز وبالتالي يجب أن تكون أدوات التنقيب عن البيانات قادرة على البحث عن أنواع مختلفة من الأنماط بأشكال متوازية لزيادة كفاءة التنقيب؛ التنقيب في البيانات هو التقاء مجموعة تخصصات (الشكل 3).



الشكل 3: التخصصات التي تلتقي في التنقيب عن البيانات

1.2.3 إجرائية التنقيب في البيانات

لإجراء التنقيب عن البيانات، يجب معالجة البيانات في بعض الخطوات لتحويلها في النهاية إلى معرفة (الشكل 4) [21].



الشكل 4: إجرائية التنقيب في البيانات

- تنظيف البيانات: يمكن أن تواجه البيانات بعض المشكلات مثل الأجزاء المفقودة أو التالفة أو غير المتوافقة. هذه المشكلة شائعة جدًا ويجب إصلاحها قبل معالجة البيانات. إن تنظيف البيانات هو عملية التخلص من البيانات السيئة أو الشاذة والاحتفاظ بالبيانات المفيدة والهادفة.
- استخراج البيانات ذات الصلة لمهمة ما: لا يمكن اعتبار جميع البيانات مفيدة لمهمة محددة، لذلك يتعين علينا فصل البيانات اللازمة لمهمتنا المحددة لاستخراج البيانات.

- تطبيق التنقيب عن البيانات: يمكن الآن تطبيق أي من خوارزميات التنقيب عن البيانات لأداء تقنية استخراج البيانات. يتم اختيار خوارزمية التنقيب وفقًا للمهمة التي نريد تنفيذها ونموذج التنقيب المستخدم.
- أنماط الاستخراج: يمكن استخدام نتائج التنقيب عن البيانات لاستخراج الأنماط المفيدة التي تمثل المعرفة المطلوبة من عملية استخراج البيانات بأكملها.

2.2.3 محاور التنقيب في البيانات Data Mining Tasks

- التنقيب في قواعد الاقتران Association Rules mining .
- التصنيف Classification .
- العنقدة clustering وأحياناً تسمى التجزئة Segmentation.

1.2.2.3 التنقيب في قواعد الاقتران

وهي ايجاد علاقات الارتباط بين الكائنات مع بعضها البعض. لنفترض أن مجموعات من الكائنات بحيث تحتوي كل مجموعة على كائنات محددة ليست بالضرورة متماثلة في مجموعات أخرى [22] ، تكمن المشكلة هنا هي العثور على علاقة وجود كائن واحد بوجود كائنات أخرى في نفس المجموعة. لذلك يمكن تحديد التنقيب في قواعد الارتباط من خلال إيجاد أنماط متكررة أو ارتباطات أو هياكل سببية بين مجموعات العناصر أو الكائنات في قواعد بيانات المعاملات وقواعد البيانات العلائقية ومستودعات المعلومات الأخرى.

تستخدم خوارزميات قواعد الاقتران من أجل اكتشاف علاقات هامة مخفية في مجموعة بيانات ضخمة، يمكن تمثيل العلاقات المكتشفة بشكل قواعد اقتران (Association rules) أو مجموعات من البنود المتكررة (frequent). يتم تحديد القواعد المفيدة وتقييمها اعتماداً على مقياسين أساسيين هما الدعم (support) والثقة (confidence)، يحدد الدعم مدى قابلية القاعدة للتطبيق على مجموعة البيانات، في حين تحدد الثقة كيفية تتكرر البنود في المناقلات (transactions).

2.2.2.3 التصنيف

بالتعريف هو مهمة إسناد الكائنات (object) إلى واحد أو أكثر من الفئات مسبقة التعريف [23].

بالنظر إلى مجموعة السجلات (مجموعة التدريب)، يحتوي كل سجل على مجموعة من السمات ، إحدى السمات هي الفئة، نبحث عن نموذج لسمة الفئة كدالة لقيم السمات الأخرى.

يتم استخدام مجموعة اختبار لتحديد دقة النموذج .عادةً ما يتم تقسيم مجموعة البيانات المقدمة إلى مجموعات تدريب واختبار، مع استخدام مجموعة التدريب لبناء النموذج ومجموعة الاختبار المستخدمة للتحقق من صحتها.

من أهم تقنيات التصنيف هي:

- شجرة القرار Decision Tree وهي تعني وجود سمة عند كل عقدة attribute تتفرع حسب قيمتها وصولاً إلى ورقة تحوي Class.
- الشبكات العصبية neural network
- آلة متجه الدعم support vector machines

3.2.2.3 العنقدة

العنقدة بالتعريف هي عملية فصل مجموعة بيانات إلى عدة مجموعات مختلفة أو عناقيد (cluster) مختلفة لها معنى أو مفيدة، والهدف من عنقدة الكائنات هو التقليل من التشابه بين الطبقات (الفئات) وتعظيم التشابه داخل الطبقة [24]، وكلما زاد التشابه ازداد الفرق بين الفئات الأخرى كلما كانت العنقدة أفضل وأكثر تمايزاً.

تعد خوارزمية k–means من أهم تقنيات العنقدة وهي تقنية عنقدة جزئية تستند إلى نموذج الأصل وتحاول إيجاد عدد يحدده المستخدم من العناقيد (k) الممثلة بواسطة مراكز ثقلها، وغالباً ما يكون مركز الثقل هو المتوسط (mean) لمجموعة من النقاط.

بالنظر إلى مجموعة من نقاط البيانات، لكل منها مجموعة من السمات ، وقياس التشابه فيما بينها ، يتم البحث عن مجموعات بحيث تكون نقاط البيانات في مجموعة واحدة أكثر تشابهًا مع بعضها البعض.

3.3 خوارزميات التصنيف والتنبؤ

إن خوارزميات التصنيف والتنبؤ هي شكل من أشكال تحليل البيانات والتي تستخلص نماذج تصف بشكل دقيق فئات وتصنيفات البيانات المهمة، بحيث يمكن لهذا التحليل أن يُساعد في فهم البيانات . [25]

1.3.3 أنواع خوارزميات التصنيف

يوجد أربعة أنواع من خوارزميات التصنيف وهي كما يلي:

1- التصنيف باستخدام خوارزميات شجرة القرار

2- التصنيف باستخدام خوارزميات الشبكات العصبية

3- التصنيف باستخدام نظرية الاحتمالات

4- التصنيف باستخدام خوارزمية الجار الأقرب

1.1.3.3 التصنيف باستخدام خوارزميات شجرة القرار

هي إحدى التقنيات الرئيسية المستخدمة في التصنيف والتنبؤ، وهي نموذج استكشافي يظهر على شكل شجرة ويمثل كل فرع من فروع الشجرة سؤالاً تصنيفياً [26]، وتمثل أوراق الشجرة أجزاء من قاعدة البيانات تنتمي للتصنيفات التي تم بنائها.

تركز على التعلم من مجموعة من الأمثلة غير المرتبة وغير النظامية. الغرض من إنشاء شجرة قرار هو إيجاد العلاقة بين السمات والفئات، واستخدامها للتنبؤ بفئات سجلات الفئات غير المعروفة في المستقبل .يستخدم أسلوبًا تنازليًا لمقارنة السمات في العقد الداخلية لشجرة القرار، والحكم على الفرع الهابط من العقدة وفقًا لقيم السمات المختلفة ، والحصول على استنتاج في العقدة الورقية لشجرة القرار. تتضمن خوارزميات شجرة القرار الرئيسية خوارزميات ID3 و C4.5 (C5.0) و CART و PUBLIC و SLIQ و SPRINT . لديها اختلافاتهم الخاصة في التكنولوجيا المستخدمة لتحديد سمات الاختبار،

وهيكل شجرة القرار التي تم إنشاؤها ، وطريقة التقليم، والقدرة على التعامل مع مجموعات البيانات الكبيرة.

إن شجرة القرار والخوارزميات التي تُستخدم لإنتاجها يمكن أن تكون معقدة ولكن النتائج التي تؤدي لها يمكن اظهارها بشكل مبسط وسهل الفهم وبفائدة عالية المستوى، وعلى الرغم من أن أشجار القرار تستخدم في الاستكشاف وتحضير البيانات للعمليات الإحصائية إلا أنها أيضاً تستخدم بشكل أكثر للتنبؤ أو التوقع.

2.1.3.3 تصنيف باستخدام خوارزميات الشبكات العصبية

تعتبر خوارزميات الشبكات العصبية Neural Netwoks أهم خوارزميات تنقيب البيانات، نظراً للنتائج التي يتم التوصل إليها باستخدامها، وكذلك لإمكانية تطبيقها لحل العديد من المشاكل وبكافة أنواعها، بالرغم من تعقيدها مقارنة بالخوارزميات الأخرى وهو ما أدى إلى عدم انتشارها بشكل أوسع [27].

تعمل خوارزميات الشبكات العصبية بنفس طريقة العقل البشري في نقل ومعالجة وتحليل المعلومات والتوصل إلى الاستنتاجات واكتشاف الأنماط والتنبؤات.

خوارزميات الشبكات العصبية هي نموذج رياضي يطبِّق بُنية مشابهة للوصلات الشبكية لأعصاب الدماغ لمعالجة المعلومات، حيث يتم توصيل عدد كبير من العقد (أو "الخلايا العصبية" أو "الوحدات") ببعضها البعض لتشكيل شبكة، أي "شبكة عصبية" لتحقيق الغرض من معالجة المعلومات. عادة ما تحتاج الشبكات العصبية إلى التدريب، وعملية التدريب هي عملية التعلم للشبكة. يغير التدريب قيمة وزن الاتصال لعقدة الشبكة لجعلها تحتوي على وظيفة تصنيف، ويمكن استخدام الشبكة المدربة للتعرف

على الكائنات. في الوقت الحاضر، هناك المئات من النماذج المختلفة للشبكات العصبية، مثل شبكة BP، وشبكة RBF ذات الأساس الشعاعي ، وشبكة Hopfield، والشبكة العصبية العشوائية (Boltzmann)، والشبكة العصبية التنافسية (شبكة Hamming)، وشبكة الخرائط ذاتية التنظيم .....إلخ. ومع ذلك، لا تزال الشبكات العصبية الحالية بشكل عام تعاني أوجه قصور مثل سرعة التقارب البطيئة، والكمية الكبيرة من العمليات الحسابية، ووقت التدريب الطويل وضعف القدرة على التفسير.

3.1.3.3 التصنيف باستخدام نظرية الاحتمالات: (Bayes Theory)

هي طريقة تصنيف إحصائية وتعتمد فكرتها على بناء الاحتمالات، وذلك من خلال التنبؤ باحتمال أن ينتمي سجل من سجلات قاعدة البيانات لفئة محددة، وتستند هذه الطريقة في التصنيف على النظرية الافتراضية أو نظرية الاحتمالات (Bayes Theory) وتعرّف باسم العالم "بايز" (Bayes) الذي وضعها في القرن الثامن عشر [28] .

تستخدم هذه الخوارزميات أساساً نظرية بايز للتنبؤ باحتمالية أن تنتمي عينة من فئة غير معروفة إلى كل فئة، وتحديد الفئة الأكثر احتمالية باعتبارها الفئة النهائية للعينة .نظرًا لأن إنشاء نظرية بايز نفسها يتطلب افتراضًا قويًا للاستقلال المشروط، وغالبًا ما يكون هذا الافتراض غير صالح في المواقف الفعلية، وبالتالي ستنخفض دقة تصنيفها. لهذا السبب، كان هناك العديد من خوارزميات تصنيف Bayesianالتي تقلل من افتراض الاستقلال، مثل خوارزمية TAN (Tree Augmented Nave Bayes)، والتي يتم تنفيذها عن طريق زيادة الارتباط بين أزواج السمات على أساس بنية شبكة .

4.1.3.3 التصنيف باستخدام خوارزمية الجار الأقرب

تعتبر خوارزمية الجار الأقرب (k–Nearest Neighbours) من خوارزميات التصنيف والتنبؤ التي تهدف للتنبؤ عن طريق مقارنة السجلات الشبيهة بالسجل المراد التنبؤ له وتقدر القيمة المجهولة لهذا السجل بناء على معلومات لتلك السجلات، بمعنى آخر طريقة تصنيف قائمة على المثيل: تتمثل الطريقة في العثور على عينات التدريب k الأقرب للعينة غير المعروفة x ، ومعرفة الفئة التي تنتمي إليها معظم عينات k ، ثم تصنيف x في تلك الفئة. أسلوب k القريب هو أسلوب التعلم البطيء، فهو يخزن العينات ولا يصنفها حتى يلزم التصنيف. إذا كانت مجموعة العينات أكثر تعقيدًا [29]، فقد تتسبب في الكثير من النفقات الحسابية ، لذلك لا يمكن تطبيقها على المواقف ذات الأداء القوي في الوقت الفعلي.

2.3.3 مصفوفة الارتباك

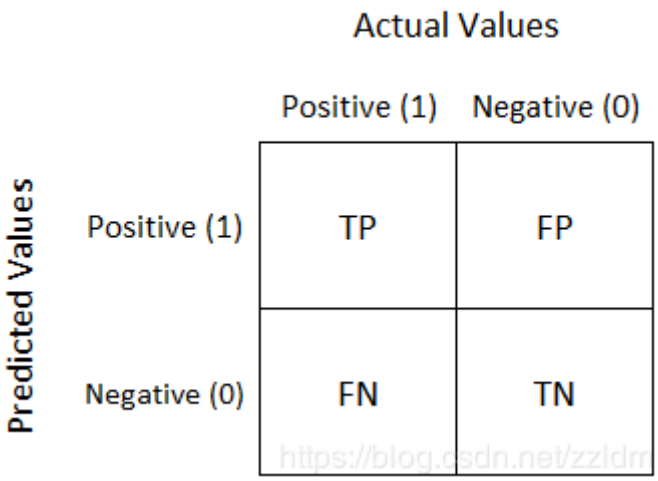
في مجال التعلم الآلي، مصفوفة الارتباك (confusion matrix) ، المعروف أيضا باسم نموذج الاحتمال أو مصفوفة خطأ، هي تخطيط جدولي معين يسمح بتصور أداء خوارزمية (الشكل 5).

يُمثل كل صف من المصفوفة حالات الفئة المتوقعة بينما يمثل كل عمود حالات الفئة الفعلية (أو العكس بالعكس).

ينبع الاسم من حقيقة أنه يجعل من السهل رؤية ما إذا كان النظام مرتبكا بين الفئتين [30] (أي يخطئ بين أحد الفئتين على أنها الأخرى).

يتم اللجوء إلى استخدام مصفوفة الارتباك لتقييم تأثير النموذج، وتكون مؤشرات تقييم النموذج في التعلم الآلي :

- الفئة الحقيقية الإيجابية True Positive (TP)
- الفئة الكاذبة الإيجابية False Positive (FP)
- الفئة الحقيقية السلبية True Negative (TN)
- الفئة الكاذبة السلبية False Negative (FN)



الشكل 5: النموذج العام لمصفوفة الارتباك

1.2.3.3 معايير التقييم

- الاستدعاء Recall وتعبر عنه المعادلة التالية:

$$Recall = \frac{TP}{TP + FN}$$

- الدقة Precision وتعبر عنه المعادلة التالية:

$$Precision = \frac{TP}{TP + FP}$$

- مقياس F Measure F−وتعبر عنه المعادلة التالية:

$$F\ Measure = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

- معدل الإيجابي الحقيقي TP Rate وتعبر عنه المعادلة التالية:

$$TPR = \frac{TP}{TP + FN}$$

- معدل الإيجابي الكاذب FP Rate وتعبر عنه المعادلة التالية:

$$FPR = \frac{FP}{FP + TN}$$

3.3.3 مثال عملي لطريقة التصنيف باستخدام نظرية الاحتمالات

بافتراض قاعدة بيانات لمبيعات الأجهزة الالكترونية [31] (الجدول 1).

الجدول 1: مجموعة بيانات لمبيعات أجهزة الكترونية

م	العمر	الدخل	طالب	التصنيف المالي	الفئة : شراء حاسب
1	شاب صغير	مرتفع	لا	متوسط	لا
2	شاب صغير	مرتفع	لا	ممتاز	لا
3	متوسط العمر	مرتفع	لا	متوسط	نعم

4	رجل كبير	متوسط	لا	متوسط	نعم
5	رجل كبير	منخفض	نعم	متوسط	نعم
6	رجل كبير	منخفض	نعم	ممتاز	لا
7	متوسط العمر	منخفض	نعم	ممتاز	نعم
8	شاب صغير	متوسط	لا	متوسط	لا
9	شاب صغير	منخفض	نعم	متوسط	نعم
10	رجل كبير	متوسط	نعم	متوسط	نعم
11	شاب صغير	متوسط	نعم	ممتاز	نعم
12	متوسط العمر	متوسط	لا	ممتاز	نعم
13	متوسط العمر	مرتفع	نعم	متوسط	نعم
14	رجل كبير	متوسط	لا	ممتاز	لا

لو كان المطلوب هو سجل ما لا ينتمي لقاعدة البيانات والمراد تصنيفه وهو :

العمر = شاب صغير ، الدخل=متوسط، طالب= نعم، التصنيف المالي = متوسط

من بيانات الجدول واعداد تكرارات السجلات في كل فئة مقارنة بالعدد الإجمالي فيها يمكن حساب القيم التالية:

احتمالات شراء جهاز حاسب أو عدم شراءه :

(ن) شراء حاسب = نعم هو 14/9 = 0.643

(ل) شراء حاسب = لا هو 14/5= 0.357

احتمالات شراء جهاز حاسب بحسب سمات الزبون المختلفة (الجدول 2).

الجدول 2: احتمالات شراء جهاز حاسب بحسب سمات الزبون المختلفة

	العمر= شاب صغير		**الدخل= متوسط**		**طالب= نعم**		**التصنيف المالي= متوسط**	
شراء حاسب	نعم	لا	نعم	لا	نعم	لا	نعم	لا
	9/2	5/3	9/4	5/2	9/6	5/1	9/6	5/2
	0.222	0.600	0.444	0.400	0.667	0.200	0.667	0.400

من كل الاحتمالات السابقة يمكن حساب احتمال (شراء حاسب=نعم) لكل السمات المطلوبة معاً للسجل الذي يتم استكشافه:

احتمال (**شراء حاسب= نعم**) = 0.222 * 0.444 * 0.667 * 0.667

= 0.0444

احتمال **(شراء حاسب=لا)** = 0.600 * 0.400 * 0.200 * 0.400

= 0.019

مع ملاحظة أنه تم ضرب جميع الاحتمالات للحصول على الاحتمال الإجمالي حيث أن جميع السمات منفصلة عن بعضها ولا تعتمد على بعضها البعض (Independent) حسب نظرية الاحتمالات.

ويكون:

(شراء حاسب= نعم) * (شراء حاسب= نعم (ن)) = 0.044 * 0.643

= 0.028

ويكون:

(شراء حاسب= لا) * (شراء حاسب= لا (ل)) = 0.019 * 0.357

= 0.007

وحيث أنه لا توجد إلا فئتين فقط (نعم، لا)، وأن احتمال الانتماء للفئة (نعم) أكبر من احتمال الانتماء للفئة (لا) فستكون النتيجة أن نموذج التصنيف باستخدام النظرية الافتراضية يتوقع النتيجة (نعم) للسجل المستكشف.

أي أن السجل بالسمات التالية: (العمر = شاب صغير ، الدخل=متوسط، ،طالب= نعم،، التصنيف المالي = متوسط)، يُتوقع له أن يقوم بشراء جهاز حاسب.

القسم العملي

الفصل الرابع: الحل المقترح

1.4 جمع ومعالجة البيانات

اعتماداً على البيانات الموجودة في وزارة الموارد المائية تم جمع المعلومات اللازمة لتشكيل مجموعة البيانات، (هذه البيانات خاصة بوزارة الموارد المائية ولا يمكن مشاركتها).

توجد هذه البيانات في ملفات مختلفة وليست في مجموعة واحدة، وبما أن مصادر البيانات مختلفة فقد تطلب معالجة البيانات أعمال إضافية لتعمل بالشكل المطلوب.

البيانات اللازمة لتنفيذ المشروع هي:

1. سعر برميل النفط.
2. نسبة الهطول المطري.
3. درجة الحرارة.
4. الموسم السنوي (الفصل في السنة).
5. وارد نهر الفرات.

تم تحويل البيانات الرقمية إلى فئات متعددة (الجداول 3،4،5) بعد المناقشة مع الخبراء المتخصصين في وزارة الموارد المائية وذلك بغرض تبسيط المسألة كما تم تحويل مسألة التصنيف لتصنيف ثنائي (الجدول 6).

الجدول 3: تحويل القيم الرقمي المتعلقة بسعر برميل النفط إلى فئات

المجال رقمياً	الفئة
60 وما فوق	Very High
26 إلى 59.9	High
10 إلى 25.9	Medium
1.5 إلى 9.9	Low
0.65 إلى 1.4	Very Low

الجدول 4: تحويل القيم الرقمي المتعلقة بدرجات الحرارة إلى فئات

المجال رقمياً	الفئة
23 وما فوق	Very High
18 إلى 22.9	High
9 إلى 17.9	Medium
1 إلى 8.9	Low

5- إلى 0.9	Very Low

الجدول 5: تحويل القيم الرقمي المتعلقة بالهطول المطري إلى فئات

المجال رقمياً	الفئة
79 وما فوق	Very High
50 إلى 78.9	High
30 إلى 49.9	Medium
8 إلى 29.9	Low
2 إلى 7.9	Very Low

الجدول 6: تحويل القيم الرقمي المتعلقة بوارد نهر الفرات إلى فئات

المجال رقمياً	الفئة
500 وما فوق	High
80 إلى 499	Low

2.4 مجموعة البيانات النهائية

بعد جمع البيانات وإجراء عمليات المعالجة عليها أصبحت مجموعة البيانات التي تتضمن 1047 عينة وتتألف من خمسة حقول –أربعة مدخلات وخرج وحيد – (الجدول 7). (الشكل 6) يوضح الصيغة النهائية لمجموعة البيانات بعد المعالجة.

الجدول 7: مجموعة البيانات بعد المعالجة

الواصفة	الشرح	القيم	النوع
Oil	سعر برميل النفط	VH H M L VL	دخل
Temp	درجة الحرارة	VH H M L VL	دخل
Rain	نسبة الهطول المطري	VH H M L VL	دخل

Season	الموسم في السنة (الفصل: على الترتيب: خريف، ربيع، صيف، شتاء)	A SP SU W	دخل
Euphrate	وارد نهر الفرات	H L	خرج

Session	Rainfall	Temperature	Oil	Euphrates
A	H	L	H	H
W	H	VL	M	H
SP	H	L	L	L

الشكل 6: الصيغة النهائية لمجموعة البيانات بعد المعالجة

3.4 الخوارزميات المقترحة

بما أن التطبيق يتطلب التصنيف لتحديد وارد نهر الفرات فقد تم اختيار خوارزميات التصنيف التي تعمل بشكل أكثر فعالية مع مسائل التصنيف الثنائي [32] .

الخوارزميات التي تم اختيارها هي:

1- Naïve Bayes [33]

2- RandomTree [34]

3- OneR [35]

وبناء على التجارب التي أجريت تم اعتماد خوارزمية Naïve Bayes لتكون المصنف الأساسي في التطبيق (لمزيد من التفاصيل حول مناقشة النتائج، يرجى قراءة الفصل السادس).

4.4 التنفيذ البرمجي

1.4.4 التقنيات المستخدمة

تم استخدام لغة Python لبناء تطبيق الويب [36] (إطار العمل المستخدم هو Django [37]) وتقنيات HTML,CSS لبناء الواجهات ، تم استخدام مكتبة Bootstrap لتنسيق الواجهات [38]. المكتبات الخاصة بلغة Python والمستخدمة في المشروع هي في (الجدول 8).

الجدول 8: المكتبات المستخدمة في بناء تطبيق الويب الخاص بالمشروع

المكتبة	الوصف
Django	إطار عمل شهير لبناء تطبيقات الويب
Numpy [39]	مكتبة متخصصة في الحوسبة العلمية بلغة البايثون، وتحتوي على تشكيلة متنوعة من الأدوات والتقنيات التي من الممكن ان تستخدمها لحل مشاكل رياضية في مجالي العلوم والهندسة
Pandas [40]	مكتبة متخصصة بإجراء ما يسمى بـData Munging ، والمقصود به هو إجراء تغييرات على بيانات أساسية غير مرتبة Raw data بحيث ينتج عن هذا التغيير تحويل البيانات الى شكل أخر يُمكن فهمه والتعامل معه.
Sklearn [41]	مكتبة رئيسية للغة برمجة Python التي تستخدم عادة في مشاريع التعلم الآلي. تركز Scikit-Learn على أدوات التعلم الآلية بما في ذلك خوارزميات الأغراض الرياضية والإحصائية والعامة التي تشكل الأساس للعديد من تقنيات التعلم الآلي.
Statsmodels [41]	مكتبة توفر صفوف ووظائف للعديد من النماذج الإحصائية المختلفة.

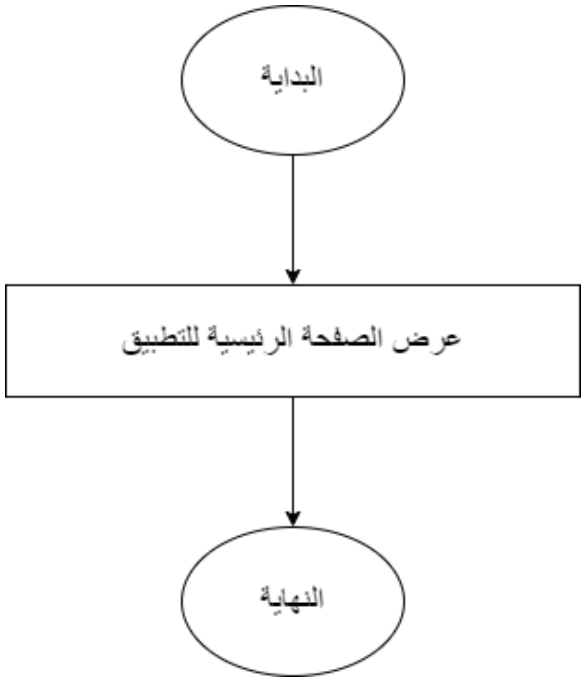
[42] Warning	مكتبة لتصفية التنبيهات من الخرج وذلك لجعله مقروء.
[43] Os	مكتبة متخصصة للتعامل مع نظام التشغيل.

2.4.4 الوظائف البرمجية

يتألف تطبيق الويب من خمسة وظائف برمجية وفق ما يلي:

1.2.4.4 التابع Main

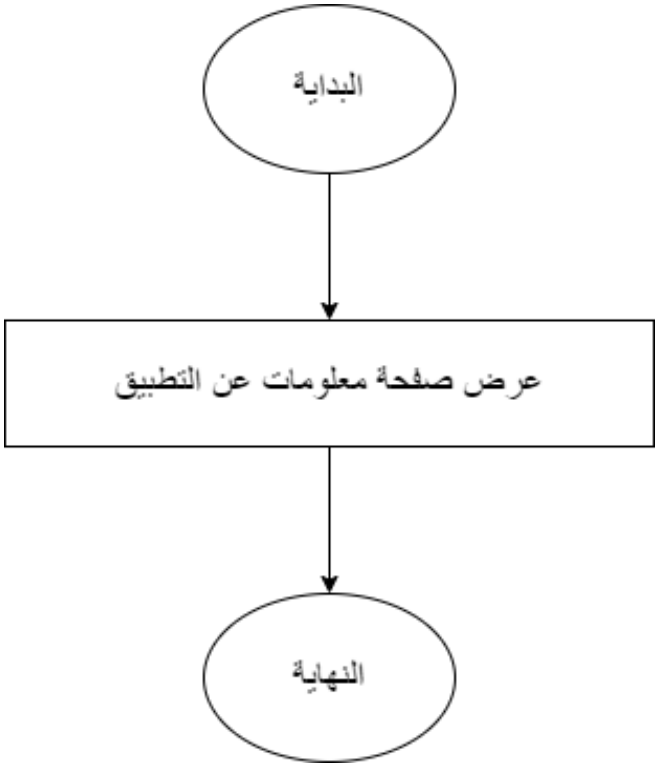
وظيفة هذا التابع عرض الصفحة الرئيسية للتطبيق والتي تتضمن المعلومات الأساسية للمشروع وأزرار الوصول للوظائف الأخرى (الشكل 7).



الشكل 7: المخطط التدفقي للتابع Main

2.2.4.4 التابع About

وظيفة هذا التابع عرض صفحة تحتوي على معلومات عن التطبيق وتتضمن هذه المعلومات أهمية سد الفرات ومساهمة التطبيق في التنبؤ بوارد النهر (الشكل 8).



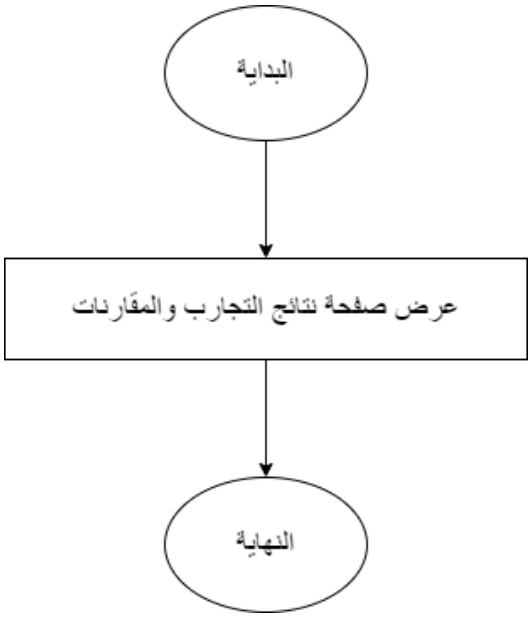
الشكل 8: المخطط التدفقي للتابع About

3.2.4.4 التابع Analysis

وظيفة هذا التابع عرض صفحة تحتوي على نتائج التجارب والمقارنات التي تم إجراؤها في مرحلة إعداد البحث (الشكل 9) وتتضمن:

1. توزع فئات كل مدخل من المدخلات بالنسبة للخرج (مخططات Histogram)
2. نتائج تدريب مجموعة البيانات التي تم إعدادها في الدراسة على ثلاثة خوارزميات تعلم آلة وهي:

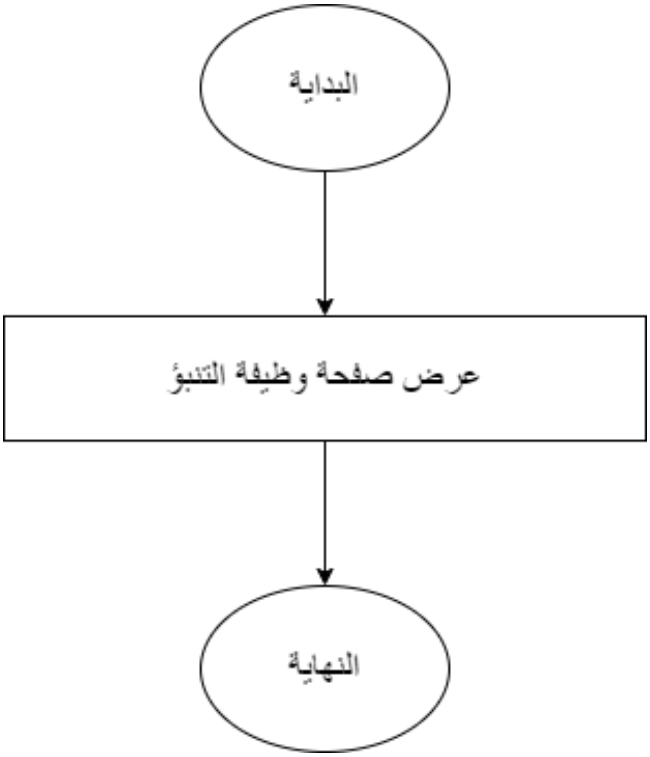
- Naïve Bayes
- OneR
- RandomTree



الشكل 9: المخطط التدفقي للتابع About

4.2.4.4 التابع Predict

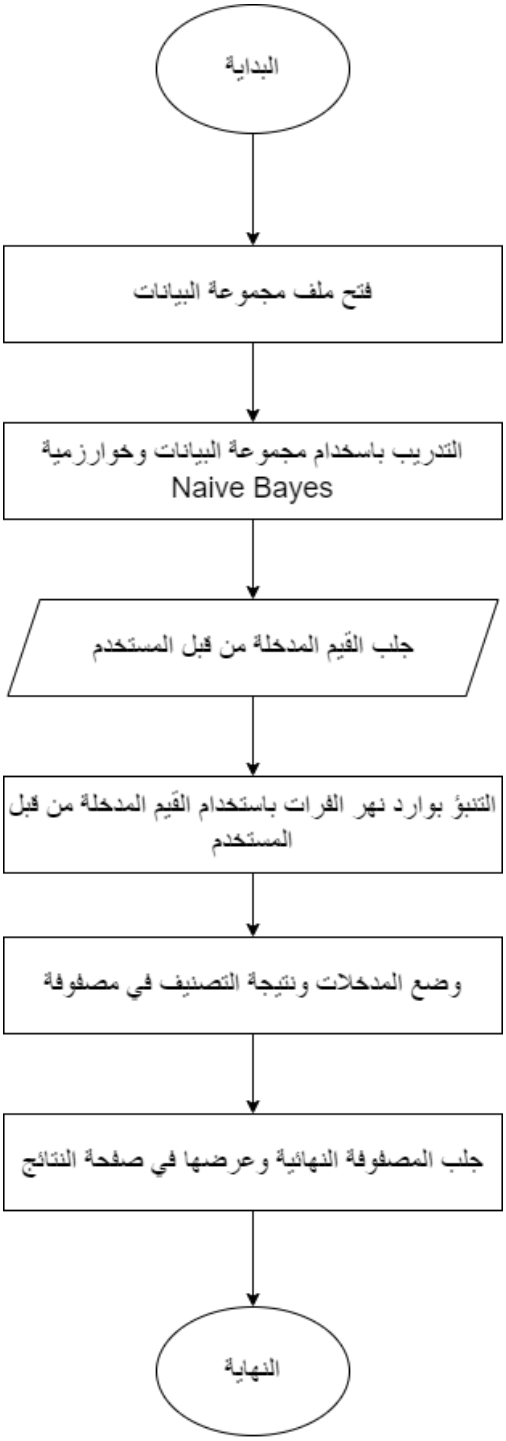
وظيفة هذا التابع عرض صفحة التطبيق التي يقوم المستخدم من خلالها إدخال القيم وتنفيذ عملية التنبؤ لمعرفة وارد نهر الفرات المتوقع (الشكل 10).



الشكل 10: المخطط التدفقي للتابع Predict

5.2.4.4 التابع Result

وهو التابع الأساسي في التطبيق ووظيفته تدريب مجموعة البيانات باستخدام خوارزمية Naïve Bayes وجلب القيم التي قام المستخدم بإدخالها وتنفيذ عملية التنبؤ لمعرفة وارد نهر الفرات وعرضه في صفحة النتائج (الشكل 11).

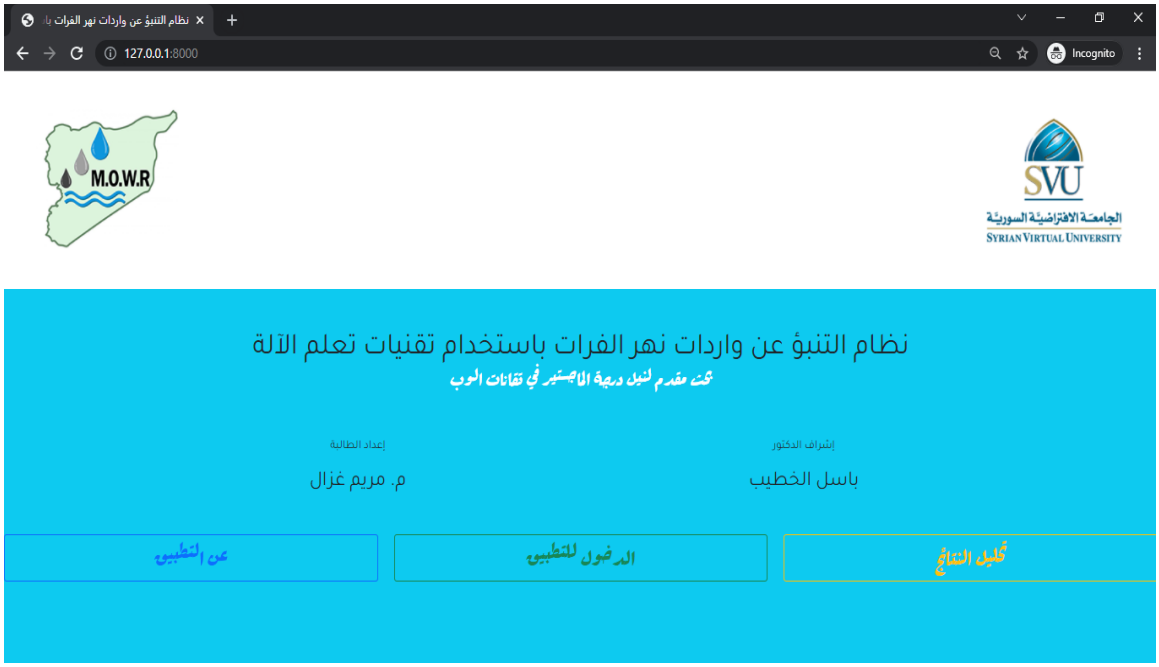


الشكل 11: المخطط التدفقي للتابع result

الفصل الخامس: دليل الاستخدام

1.5 الواجهة الرئيسية

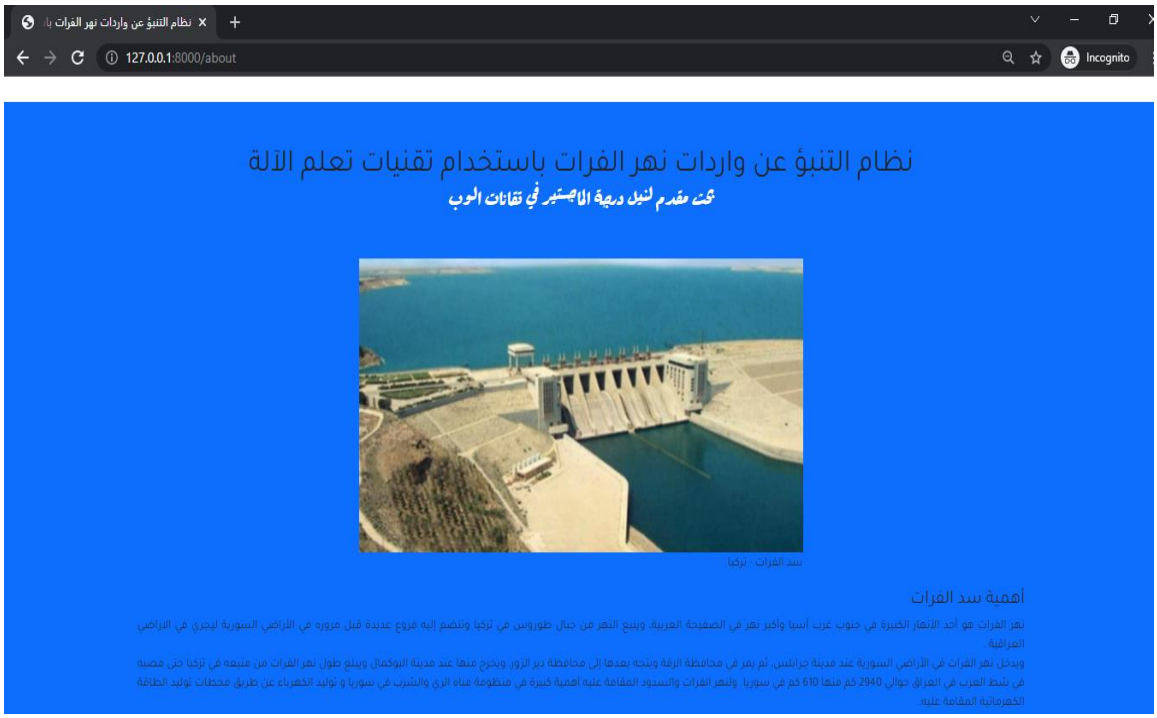
وهي الواجهة التي تظهر للمستخدم عند فتح التطبيق مباشرة وتتضمن المعلومات الأساسية عن التطبيق وأزرار الوصول للوظائف الأخرى (الشكل 12).



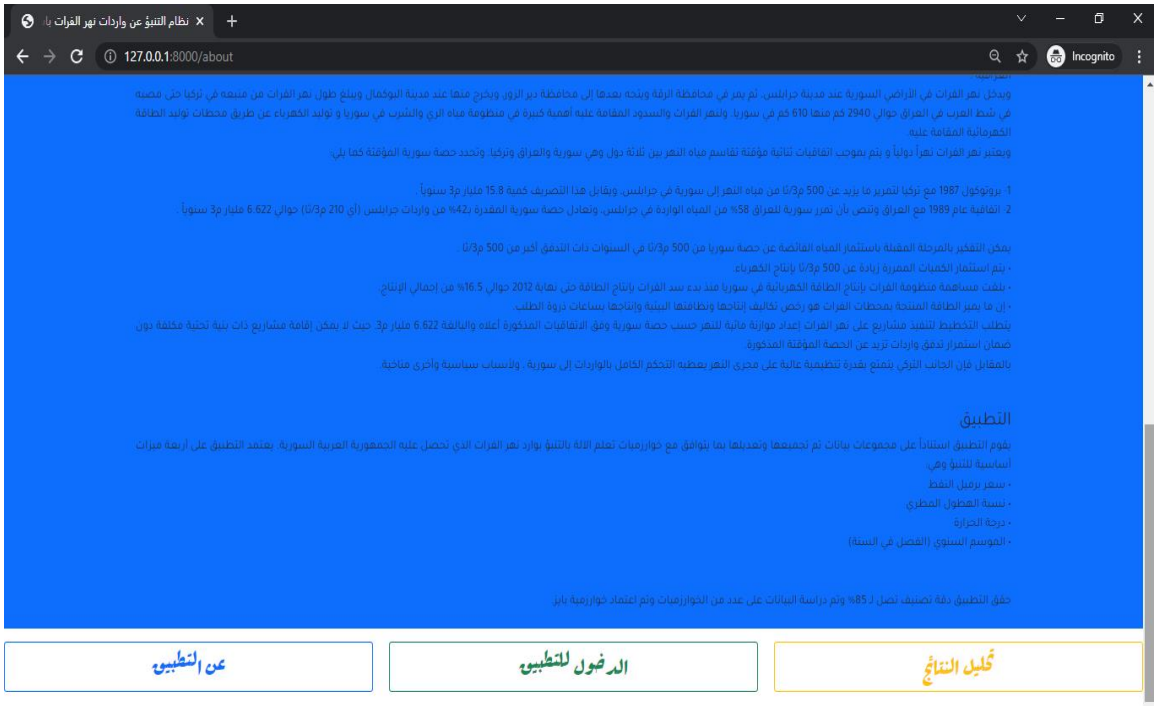
الشكل 12: الواجهة الرئيسية للتطبيق

2.5 واجهة عن التطبيق

يمكن الوصول لهذه الصفحة من الواجهة الرئيسية للتطبيق وتتضمن مقالة تعريفية عن أهمية نهر الفرات وفقرة توضح مساهمة التطبيق في التنبؤ بوارد نهر الفرات (الأشكال 13، 14). تتضمن هذه الواجهة أزرار للتنقل عبر كافة واجهات التطبيق.



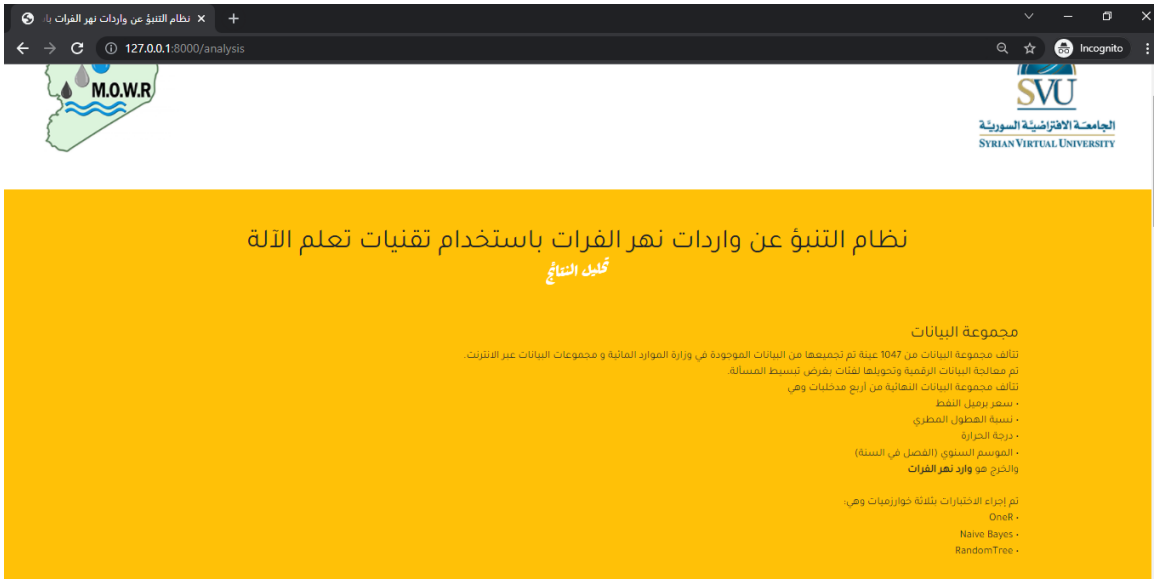
الشكل 13: واجهة معلومات التطبيق 1



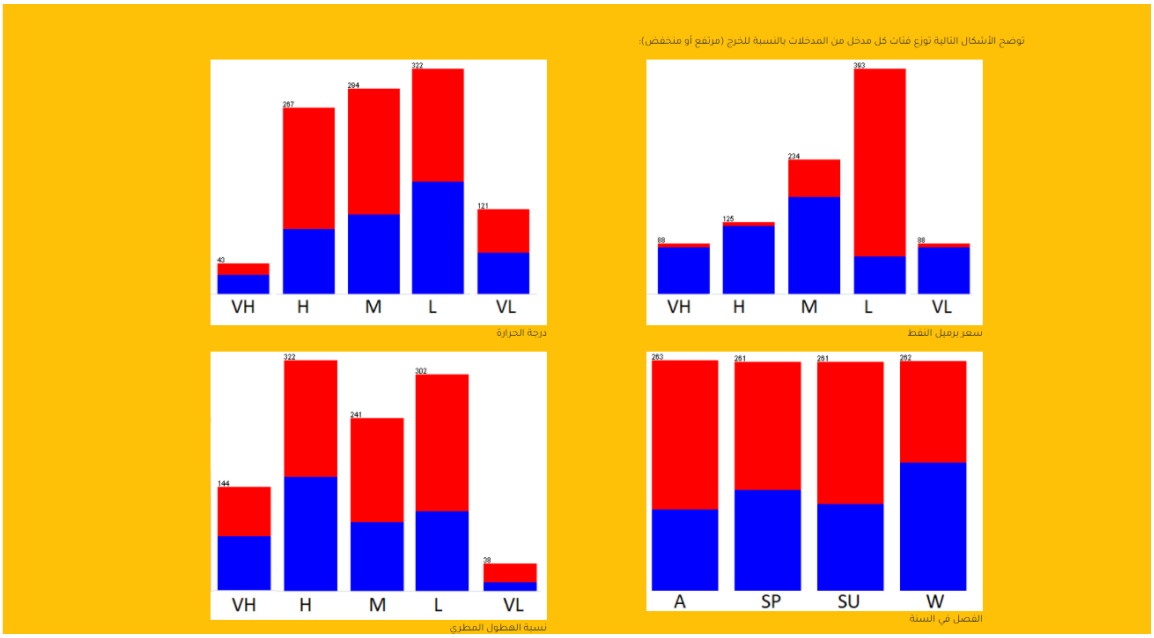
الشكل 14: واجهة معلومات التطبيق 2

3.5 واجهة تحليل النتائج

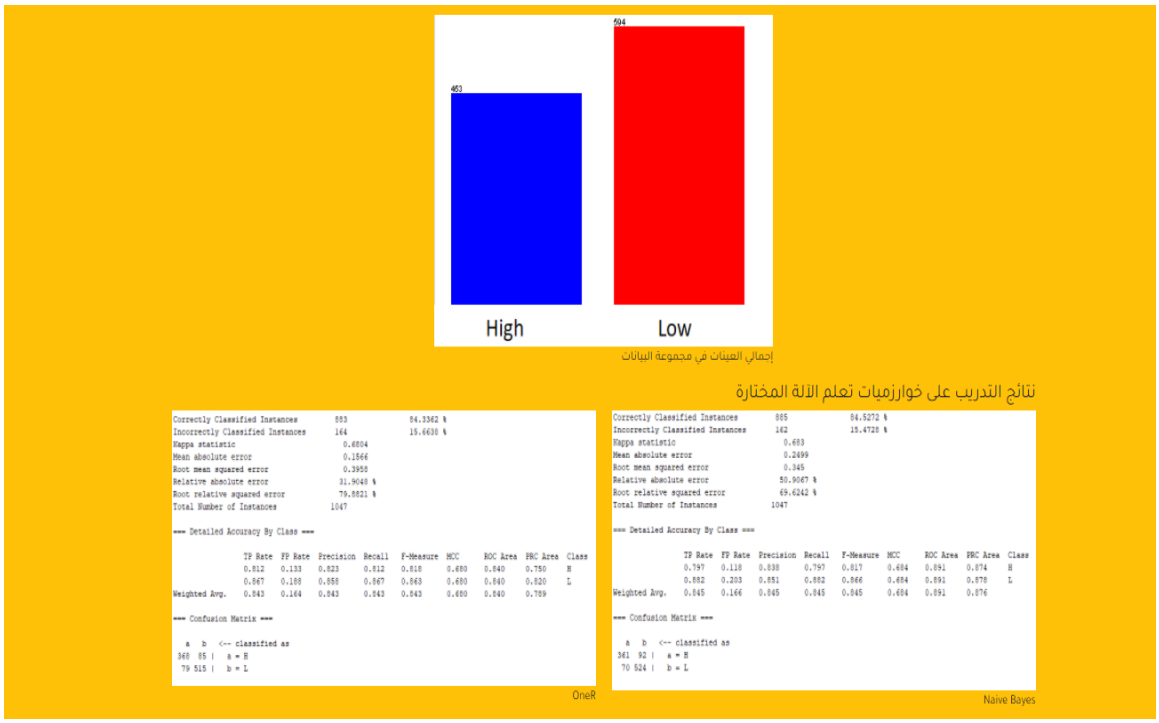
يمكن الوصول لهذه الصفحة من الواجهة الرئيسية للتطبيق وتتضمن نتائج التدريب والتجارب التي تم إجراؤها في مرحلة الدراسة وهي عرض مجموعة البيانات بشكل بياني وجداول نتائج التدريب لخوارزميات التصنيف التي تم اختيارها (الأشكال 15،16،17،18). تتضمن هذه الواجهة أزرار للتنقل عبر كافة واجهات التطبيق.



الشكل 15: واجهة تحليل النتائج 1



الشكل 16: واجهة تحليل النتائج 2

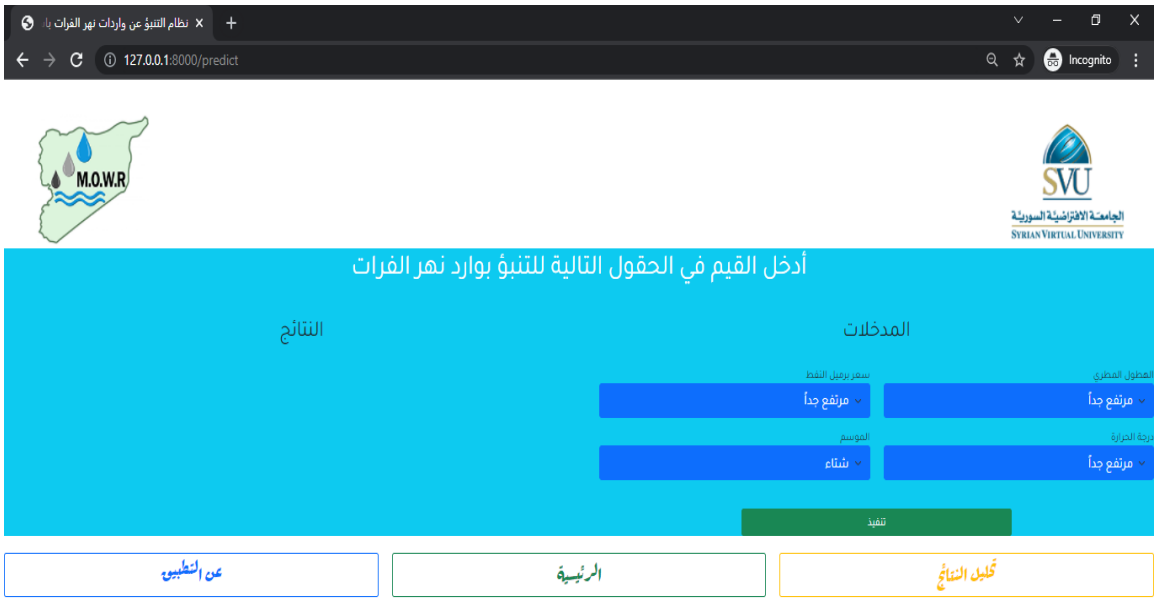


الشكل 17: واجهة تحليل النتائج 3

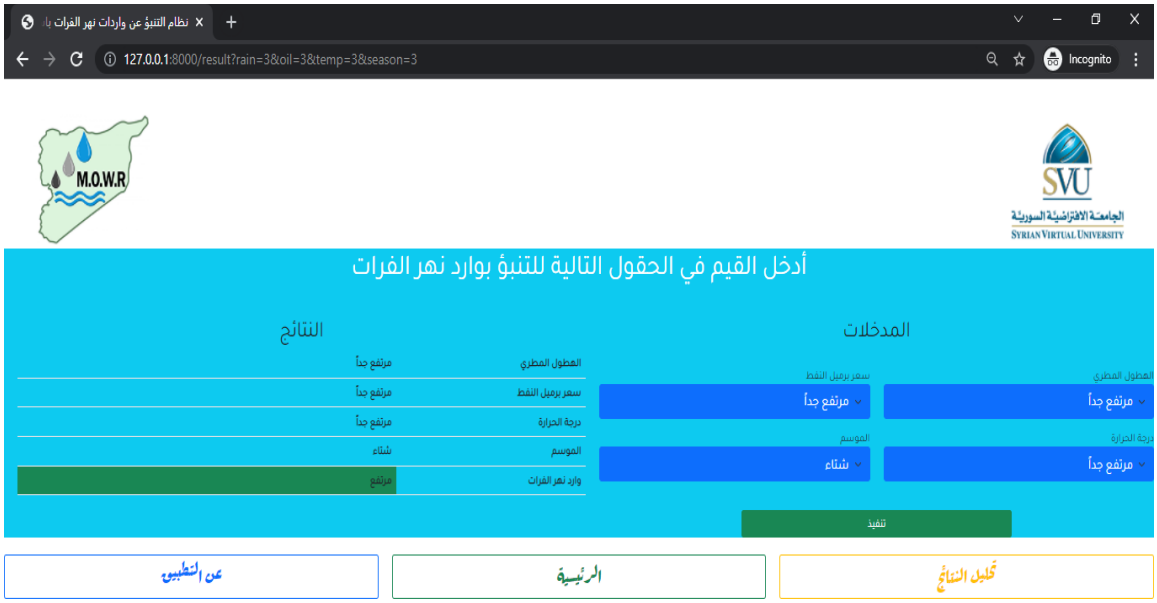
الشكل 18: واجهة تحليل النتائج 4

4.5 واجهة التنبؤ

يمكن الوصول لهذه الصفحة من الواجهة الرئيسية للتطبيق وتتضمن التابع البرمجي الذي يقوم بأخذ مدخلات المستخدم والتنبؤ بوارد نهر الفرات استناداً عليها. (الأشكال 19،20). تتضمن هذه الواجهة أزرار للتنقل عبر كافة واجهات التطبيق.



الشكل 19: واجهة التنبؤ قبل التنفيذ



الشكل 20: واجهة التنبؤ بعد التنفيذ

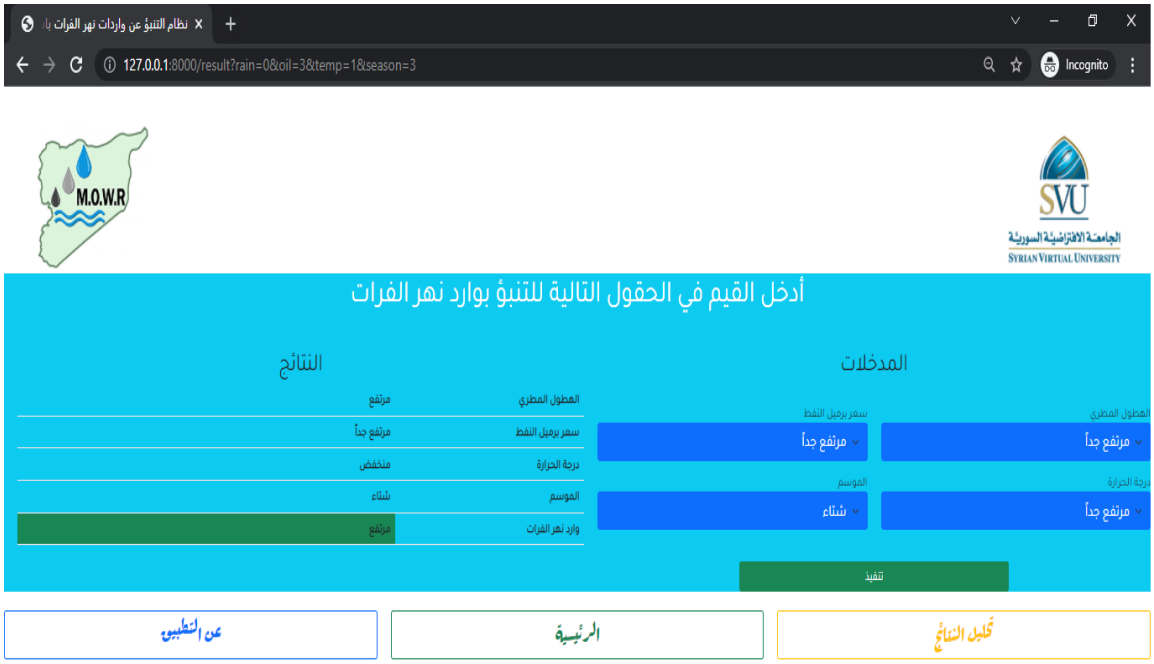
5.5 أمثلة عملية لاستخدام التطبيق

1.5.5 المثال الأول

سنقوم بإدخال المدخلات التالية (الجدول 9) وتكون النتيجة هي وارد مرتفع لنهر الفرات (الشكل 21).

الجدول 9: مدخلات المثال الأول

سعر برميل النفط	مرتفع جداً
الهطول المطري	مرتفع
درجة الحرارة	منخفض
الموسم	شتاء



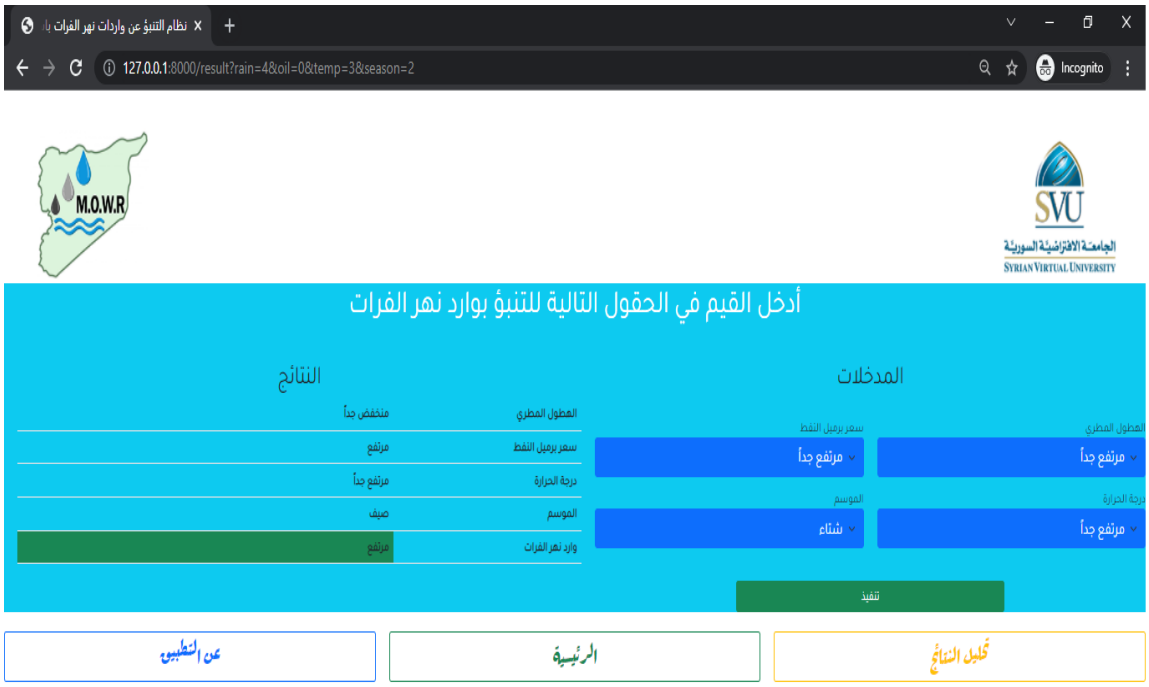
الشكل 21: وارد نهر الفرات بناءً على مدخلات المثال الأول

2.5.5 المثال الثاني

بالعمل على إدخال المدخلات التالية (الجدول 10) وتكون النتيجة هي وارد مرتفع لنهر الفرات (الشكل 22).

الجدول 10: مدخلات المثال الثاني

سعر برميل النفط	مرتفع
الهطول المطري	منخفض جداً
درجة الحرارة	مرتفع جداً
الموسم	صيف



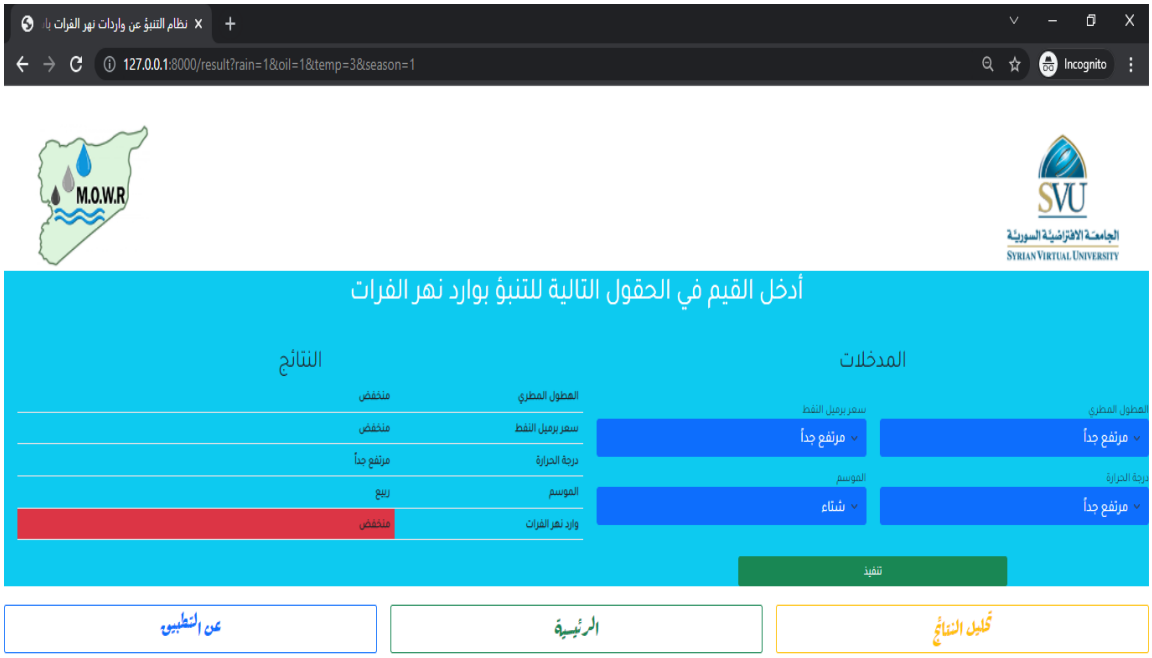
الشكل 22: وارد نهر الفرات بناءً على مدخلات المثال الثاني

3.5.5 المثال الثالث

بالعمل على إدخال المدخلات التالية (الجدول 11) وتكون النتيجة هي وارد منخفض لنهر الفرات (الشكل 23).

الجدول 11: مدخلات المثال الثالث

سعر برميل النفط	منخفض
الهطول المطري	منخفض
درجة الحرارة	مرتفع جداً
الموسم	ربيع



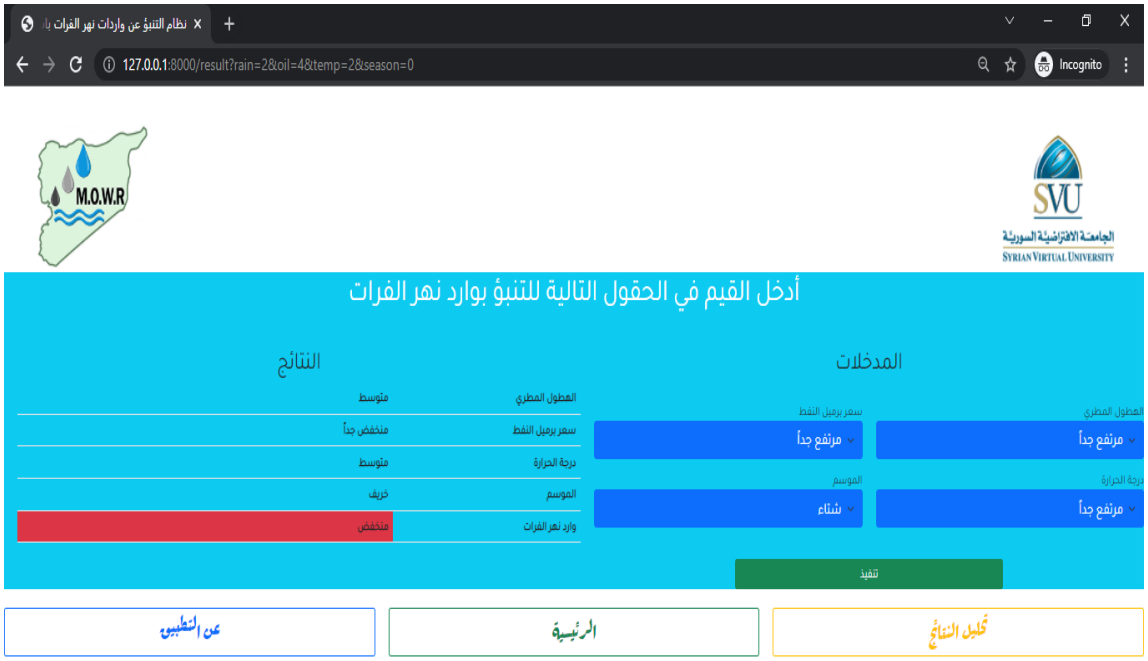
الشكل 23: وارد نهر الفرات بناءً على مدخلات المثال الثالث

4.5.5 المثال الرابع

بالعمل على إدخال المدخلات التالية (الجدول 12) وتكون النتيجة هي وارد منخفض لنهر الفرات (الشكل 24).

الجدول 12: مدخلات المثال الرابع

سعر برميل النفط	مرتفع جداً
الهطول المطري	مرتفع
درجة الحرارة	منخفض
الموسم	شتاء

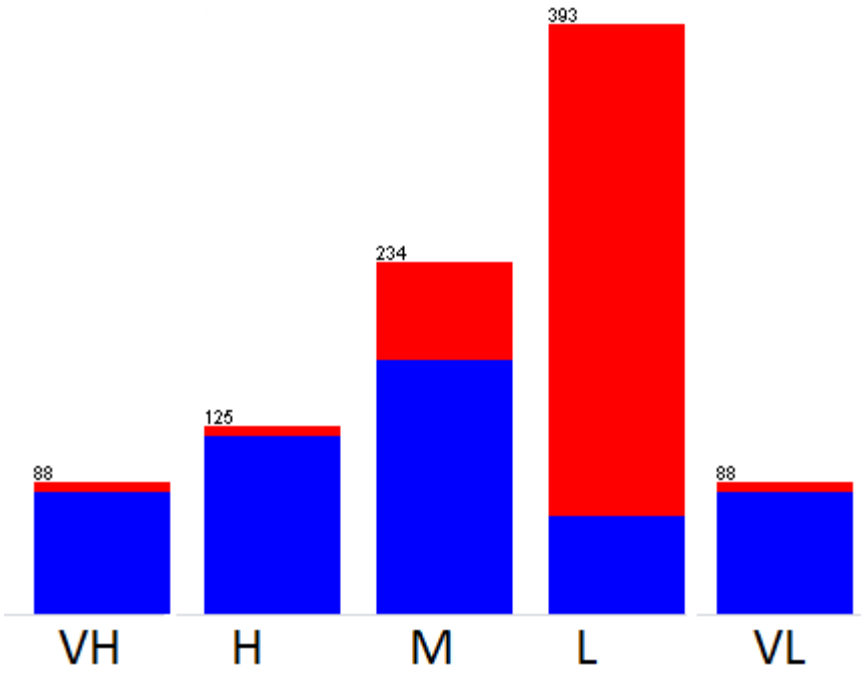


الشكل 24:وارد نهر الفرات بناءً على مدخلات المثال الرابع

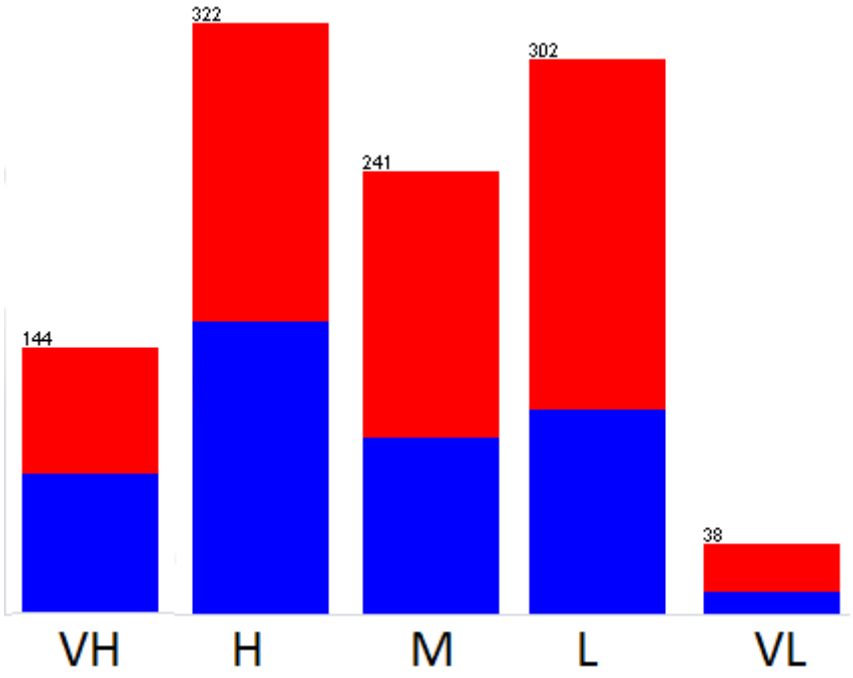
الفصل السادس: النتائج والمقارنات

1.5 تحليل المدخلات بيانياً

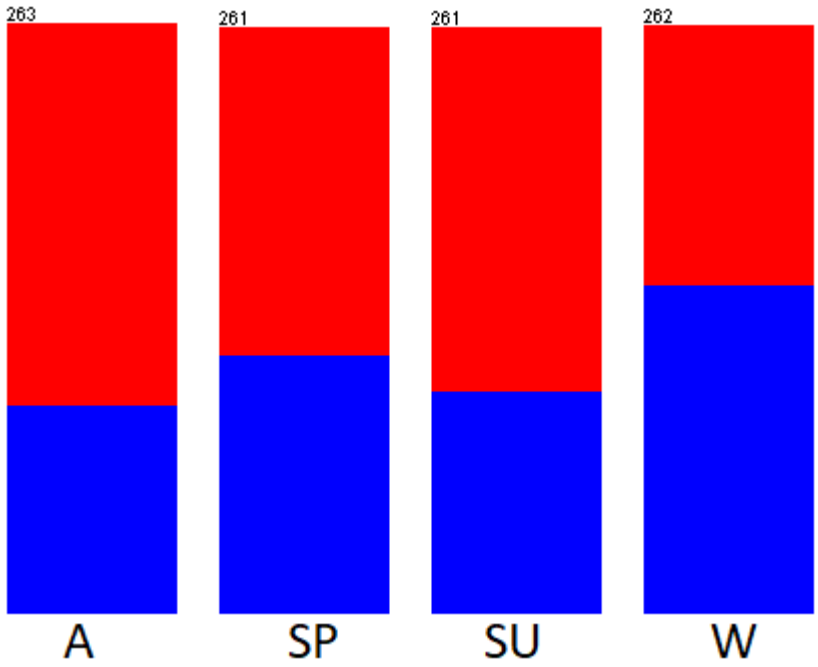
تتألف مجموعة البيانات النهائية من أربعة مدخلات وخرج وحيد، لتصور العلاقات بين هذه الأعمدة تم تمثيلها بيانياً بواسطة برنامج WEKA [44] الذي تم استخدامه في مرحلة التجارب (الأشكال 25،26،27،28،29)، يمثل اللون الأزرق العينات التي يكون فيها وارد نهر الفرات مرتفع، واللون الأحمر يمثل العينات التي يكون فيها وارد نهر الفرات منخفض.



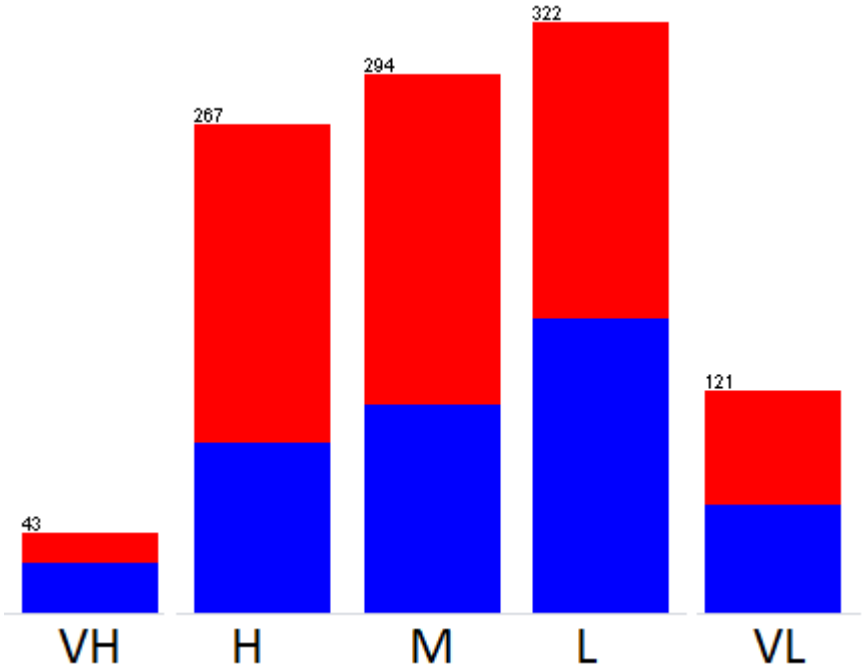
الشكل 25: فئات مدخل "سعر برميل النفط" بالنسبة لعينات مجموعة البيانات (وارد نهر الفرات)



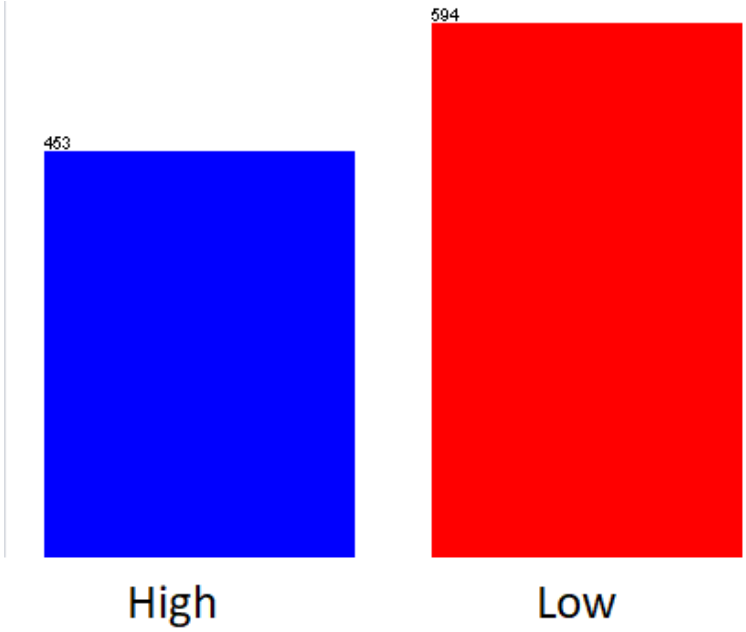
الشكل 26: فئات مدخل "الهطول المطري" بالنسبة لعينات مجموعة البيانات (وارد نهر الفرات)



الشكل 27: فئات مدخل "الموسم في السنة" بالنسبة لعينات مجموعة البيانات (وارد نهر الفرات)



الشكل 28: فئات مدخل "درجة الحرارة" بالنسبة لعينات مجموعة البيانات (وارد نهر الفرات)



الشكل 29: فئات وارد نهر الفرات بالنسبة لعدد العينات في مجموعة البيانات

2.5 اختيار الخوارزميات المناسبة

بعد انتهاء عمليات معالجة وجمع البيانات أصبحت مجموعة البيانات النهائية تتضمن أربعة مدخلات وخرج وحيد، وكل هذه الحقول هي فئات.

بما أن مسألة التصنيف هي مسألة تصنيف ثنائي فقد تم اختيار الخوارزميات التالية: Naïve Bayes ، OneR ، Random Tree .

هذه الخوارزميات تعمل بكفاءة مع البيانات المصنفة كفئات ومع التصنيف الثنائي بشكل خاص.

3.5 التدريب

تم استخدام برنامج WEKA في مرحلة التدريب، كما تم استخدام ثلاثة طرق للتدريب وتقييم نتائجه وهي:

1. طريقة النسب 80/20: وفي هذه الطريقة تم تقسيم مجموعة البيانات إلى 20% كمجموعة اختبار و80% كمجموعة تدريب.
2. طريقة النسب 66/34: وفي هذه الطريقة تم تقسيم مجموعة البيانات إلى 34% كمجموعة اختبار و66% كمجموعة تدريب.
3. طريقة Cross-validation: وفي هذه الطريقة يتم تقسيم مجموعة البيانات إلى مجموعة اختبار ومجموعة تدريب عدد من المرات وبشكل عشوائي ومختلف استناداً لعامل يسمى Folds، تم تحديد العامل في هذه الطريقة بالقيمة 10. تستخدم هذه الطريقة للتحقق من عدم وجود overfitting [46].

1.3.5 التدريب باستخدام خوارزمية OneR

1.1.3.5 الطريقة الأولى

حقق النموذج دقة تصنيف 86.12% ، (الجدول 13) يوضح نتائج العملية كاملة.

الجدول13 : تدريب مجموعة البيانات باستخدام خوارزمية OneR بالطريقة الأولى

Summary		
Correctly Classified Instances	180	86.12 %
Incorrectly Classified Instances	29	13.87 %
Kappa statistic	0.719	
Mean absolute error	0.138	
Root mean squared error	0.3725	
Relative absolute error	28.01 %	
Root relative squared error	74.34 %	
Total Number of Instances	209	

Detailed Accuracy By Class					
	TP Rate	FP Rate	Precision	Recall	F-Measure
	0.806	0.090	0.888	0.806	0.845
	0.910	0.194	0.842	0.910	0.874
Weighted Avg	0.861	0.145	0.863	0.861	0.861

Confusion Matrix		
H	L	
79	19	H
10	101	L

2.1.3.5 الطريقة الثانية

حقق النموذج دقة تصنيف 84.55% ، (الجدول 14) يوضح نتائج العملية كاملة.

الجدول14 : تدريب مجموعة البيانات باستخدام خوارزمية OneR بالطريقة الثانية

Summary					
Correctly Classified Instances		301		84.55 %	
Incorrectly Classified Instances		55		15.44 %	
Kappa statistic		0.68			
Mean absolute error		0.15			
Root mean squared error		0.39			
Relative absolute error		31.09%			
Root relative squared error		77.75%			
Total Number of Instances		356			
Detailed Accuracy By Class					
	TP Rate	FP Rate	Precision	Recall	F-Measure
	0.797	0.109	0.873	0.797	0.833
	0.891	0.203	0.824	0.891	0.856
Weighted Avg	0.846	0.158	0.848	0.846	0.845
Confusion Matrix					
H		L			
37		35		H	
20		164		L	

3.1.3.5 الطريقة الثالثة

حقق النموذج دقة تصنيف 84.33% ، (الجدول 15) يوضح نتائج العملية كاملة.

الجدول15 : تدريب مجموعة البيانات باستخدام خوارزمية OneR بالطريقة الثالثة

Summary		
Correctly Classified Instances	883	84.33 %
Incorrectly Classified Instances	164	15.66 %
Kappa statistic	0.680	
Mean absolute error	0.156	
Root mean squared error	0.395	
Relative absolute error	31.90%	
Root relative squared error	79.88%	
Total Number of Instances	1047	

Detailed Accuracy By Class					
	TP Rate	FP Rate	Precision	Recall	F-Measure
	0.812	0.133	0.823	0.812	0.818
	0.867	0.188	0.858	0.867	0.863
Weighted Avg	0.843	0.164	0.843	0.843	0.843

Confusion Matrix		
H	L	
368	85	H
79	515	L

2.3.5 التدريب باستخدام خوارزمية RandomTree

1.2.3.5 الطريقة الأولى

حقق النموذج دقة تصنيف 81.81% ، (الجدول 16) يوضح نتائج العملية كاملة.

الجدول16 : تدريب مجموعة البيانات باستخدام خوارزمية RandomTree بالطريقة الأولى

Summary					
Correctly Classified Instances	171	81.81 %			
Incorrectly Classified Instances	38	18.18 %			
Kappa statistic	0.6319				
Mean absolute error	0.2028				
Root mean squared error	0.3455				
Relative absolute error	40.94%				
Root relative squared error	68.96%				
Total Number of Instances	209				
Detailed Accuracy By Class					
	TP Rate	FP Rate	Precision	Recall	F-Measure
	0.735	0.108	0.857	0.735	0.791
	0.892	0.265	0.792	0.892	0.839
Weighted Avg	0.818	0.192	0.823	0.818	0.817
Confusion Matrix					
H	L				
72	26	H			
12	99	L			

2.2.3.5 الطريقة الثانية

حقق النموذج دقة تصنيف 79.77% ، (الجدول 17) يوضح نتائج العملية كاملة.

الجدول17 : تدريب مجموعة البيانات باستخدام خوارزمية RandomTree بالطريقة الثانية

Summary					
Correctly Classified Instances		284	79.77 %		
Incorrectly Classified Instances		72	20.22 %		
Kappa statistic		0.592			
Mean absolute error		0.223			
Root mean squared error		0.390			
Relative absolute error		45.03%			
Root relative squared error		77.18%			
Total Number of Instances		356			
Detailed Accuracy By Class					
	TP Rate	FP Rate	Precision	Recall	F-Measure
	0.692	0.103	0.862	0.683	0.768
	0.897	0.308	0.757	0.897	0.821
Weighted Avg	0.798	0.209	0.808	0.798	0.795
Confusion Matrix					
H		L			
119		53	H		
19		165	L		

3.2.3.5 الطريقة الثالثة

حقق النموذج دقة تصنيف 83.85% ، (الجدول 18) يوضح نتائج العملية كاملة.

الجدول18 : تدريب مجموعة البيانات باستخدام خوارزمية RandomTree بالطريقة الثالثة

Summary		
Correctly Classified Instances	878	83.85 %
Incorrectly Classified Instances	169	16.14 %
Kappa statistic	0.667	
Mean absolute error	0.206	
Root mean squared error	0.348	
Relative absolute error	42.01%	
Root relative squared error	70.29%	
Total Number of Instances	1047	

Detailed Accuracy By Class					
	TP Rate	FP Rate	Precision	Recall	F-Measure
	0.764	0.104	0.848	0.764	0.804
	0.896	0.236	0.833	0.896	0.863
Weighted Avg	0.839	0.179	0.839	0.839	0.837

Confusion Matrix		
H	L	
346	107	H
62	532	L

3.3.5 التدريب باستخدام خوارزمية Naïve Bayes

1.3.3.5 الطريقة الأولى

حقق النموذج دقة تصنيف 83.73% ، (الجدول 19) يوضح نتائج العملية كاملة.

الجدول19 : تدريب مجموعة البيانات باستخدام خوارزمية Naive Bayes بالطريقة الأولى

Summary					
Correctly Classified Instances		175		83.73 %	
Incorrectly Classified Instances		34		16.26 %	
Kappa statistic		0.669			
Mean absolute error		0.248			
Root mean squared error		0.347			
Relative absolute error		50.24%			
Root relative squared error		69.32%			
Total Number of Instances		209			
Detailed Accuracy By Class					
	TP Rate	FP Rate	Precision	Recall	F-Measure
	0.735	0.072	0.900	0.735	0.809
	0.928	0.265	0.798	0.928	0.858
Weighted Avg	0.837	0.175	0.846	0.837	0.835
Confusion Matrix					
H		L			
72		26		H	
8		103		L	

2.3.3.5 الطريقة الثانية

حقق النموذج دقة تصنيف 82.86% ، (الجدول 20) يوضح نتائج العملية كاملة.

الجدول20 : تدريب مجموعة البيانات باستخدام خوارزمية Naive Bayes بالطريقة الثانية

Summary		
Correctly Classified Instances	295	82.86 %
Incorrectly Classified Instances	61	17.13 %
Kappa statistic	0.654	
Mean absolute error	0.261	
Root mean squared error	0.360	
Relative absolute error	52.59%	
Root relative squared error	71.30%	
Total Number of Instances	356	

Detailed Accuracy By Class					
	TP Rate	FP Rate	Precision	Recall	F-Measure
	0.715	0.065	0.911	0.715	0.801
	0.935	0.285	0.778	0.935	0.849
Weighted Avg	0.829	0.179	0.842	0.829	0.826

Confusion Matrix		
H	L	
123	49	H
12	172	L

3.3.3.5 الطريقة الثالثة

حقق النموذج دقة تصنيف 84.52% ، (الجدول 21) يوضح نتائج العملية كاملة.

الجدول21 : تدريب مجموعة البيانات باستخدام خوارزمية Naive Bayes بالطريقة الثالثة

Summary		
Correctly Classified Instances	885	84.52 %
Incorrectly Classified Instances	162	15.47 %
Kappa statistic	0.683	
Mean absolute error	0.249	
Root mean squared error	0.345	
Relative absolute error	50.90%	
Root relative squared error	69.62%	
Total Number of Instances	1047	

Detailed Accuracy By Class					
	TP Rate	FP Rate	Precision	Recall	F-Measure
	0.797	0.118	0.838	0.797	0.817
	0.882	0.203	0.851	0.882	0.866
Weighted Avg	0.845	0.166	0.845	0.845	0.845

Confusion Matrix		
H	L	
361	92	H
70	524	L

4.5 مناقشة النتائج

يعد دقة التصنيف واحد من أهم المعايير لتقييم واختيار الخوارزمية التي سيتم اعتمادها في التطبيق، ولكن هناك عوامل أخرى لا تقل أهمية عنها وهي: سرعة التنفيذ، الكفاءة مع مسائل التصنيف الثنائي، والعمل مستقبلاً مع بيانات أضخم، والكفاءة مع الطريقة الثالثة (وهو العامل الأهم في التقييم في هذا المشروع). (الجدول 22) يوضح خلاصة عمليات التدريب التي تم إجراؤها.

الجدول 22: خلاصة عمليات التدريب باستخدام الخوارزميات المختارة بثلاثة طرق مختلفة

	Naïve Bayes	RandomTree	OneR
الطريقة الأولى	83.73%	81.81%	**86.12%**
الطريقة الثانية	82.86%	79.77%	**84.55%**
الطريقة الثالثة	**84.52%**	83.85%	84.33%

أدت خوارزمية OneR أداء أفضل في الطرق التقليدية (الأولى والثانية) وكانت خوارزمية RandomTree هي ذات الأداء الأسوء.

قدمت خوارزمية Bayes أفضل أداء في الطريقة الثالثة وهذا يعني أنها تعلمت اتخاذ القرار بشكل صحيح. ولهذا تم اختيارها للعمل في التطبيق وخاصة أنها تعمل بكفاءة عالية مع مسائل التصنيف الثنائي.

الفصل السادس: الخاتمة والأفاق المستقبلية

1.6 الخاتمة

في هذا المشروع والذي يهدف للتنبؤ بوارد نهر الفرات اعتماداً على المعطيات الحالية. تم جمع البيانات الموزعة في وزارة الموارد المائية ومعالجتها لتصبح قابلة للعمل مع خوارزميات تعلم الآلة. هذه البيانات تضمنت سعر برميل النفط والهطول المطري ودرجة الحرارة والموسم في السنة وهي المدخلات التي قامت خوارزمية تعلم الآلة المعتمدة بالتنبؤ بوارد نهر الفرات استناداً عليها. تم صياغة كافة الأعمدة في مجموعة البيانات النهائية لتصبح فئات وتم تبسيط المسألة لتصبح مسألة تصنيف ثنائي.

تم تقييم مجموعة البيانات النهائية باستخدام ثلاثة خوارزميات تصنيف شهيرة (Naïve Bayes, OneR, RandomTree) ، باستخدام ثلاثة طرق مختلفة (نسبة 80/20، نسبة 66/34 و cross validation). وحققت مجموعة البيانات دقة تصنيف تتراوح بين 79% و 86% وتم اعتماد خوارزمية Naïve Bayes التي حققت أداء أفضل مقارنة بالخوارزميتين OneR و RandomTree.

تم تنفيذ الخوارزمية برمجياً باستخدام لغة Python ضمن تطبيق ويب يتضمن وظيفة التنبؤ استناداً على المدخلات التي يتم تمريرها من قبل المستخدم وتم إضافة صفحات تعريفية تتضمن معلومات التحليل والتطبيق.

2.6 الآفاق المستقبلية

نتطلع إلى تطوير هذا العمل بإضافات هامة أبرزها:

1. إضافة عوامل جديدة تساهم بالتنبؤ ورفع دقة التصنيف.
2. التدريب باستخدام مجموعات بيانات أكبر.
3. توسيع التطبيق ليصبح منظومة كاملة تستخدم تعلم الآلة بوظائف مختلفة ويتم استخدامها من قبل الإدارة العليا في وزارة الموارد المائية.

المراجع

[1] Loucks, D. P. (2000). Sustainable water resources management. Water international, 25(1), 3-10.

[2] Biswas, A. K. (2004). Integrated water resources management: a reassessment: a water forum contribution. Water international, 29(2), 248-256.

[3] Vijay, S., & Kamaraj, K. (2019). Ground water quality prediction using machine learning algorithms in R. International Journal of Research and Analytical Reviews, 6(1), 743-749.

[4] Ji, Y., Lei, X., Cai, S., & Wang, X. (2016). Application of a classifier based on data mining techniques in water supply operation. Water, 8(12), 599.

[5] Aldhyani, T. H., Al-Yaari, M., Alkahtani, H., & Maashi, M. (2020). Water quality prediction using artificial intelligence algorithms. Applied Bionics and Biomechanics, 2020.

[6] Gakii, C., & Jepkoech, J. (2019). A classification model for water quality analysis using decision tree.

[7] Marinósdóttir, H. (2019). Applications of different machine learning methods for water level predictions (Doctoral dissertation).

[8] Salah, H. A., Mocanu, M., & Florea, A. (2014, September). Towards an integrated decision support system for the evaluation of water pollution in Tigris Basin (DSSWAPIT). In 2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 391-398). IEEE.

[9] Sun, A. Y., & Scanlon, B. R. (2019). How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. Environmental Research Letters, 14(7), 073001.

[10] Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. Water, 10(11), 1536.

[11] Sasikala, R.,. (2017). Comparative analysis of smart water resource using data mining tools. Granthaalayah, International Journal of Research, 7, 128325-128338.

[12] Humaid, E. H. (2012). A data mining based fraud detection model for water consumption billing system in MOG.

[13] Duncan, A., Chen, A. S., Keedwell, E., Djordjevic, S., & Savic, D. (2013). RAPIDS: Early warning system for urban flooding and water quality hazards.

[14] Rosegrant, M. W. (1997). Water resources in the twenty-first century: Challenges and implications for action (Vol. 20). Intl Food Policy Res Inst.

[15] دراسة خاصة بوزارة الموارد المائية – الجمهورية العربية السورية

[16] دراسة خاصة بوزارة الموارد المائية – الجمهورية العربية السورية

[17] دراسة خاصة بوزارة الموارد المائية – الجمهورية العربية السورية

[18] Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. International Journal of Information Management, 48, 63-71.

[19] Romão, W., Freitas, A. A., & Gimenes, I. M. D. S. (2004). Discovering interesting knowledge from a science and technology database with a genetic algorithm. Applied soft computing, 4(2), 121-137.

[20] Hand, D. J. (2007). Principles of data mining. Drug safety, 30(7), 621-622.

[21] Scott, P. D., & Wilkins, E. (1999). Evaluating data mining procedures: techniques for generating artificial data sets. Information and software technology, 41(9), 579-587.

[22] Ghafari, S. M., & Tjortjis, C. (2019). A survey on association rules mining using heuristics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1307.

[23] Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160(1), 3-24.

[24] Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. IEEE Transactions on neural networks, 16(3), 645-678.

[25] Kumar, R., & Verma, R. (2012). Classification algorithms for data mining: A survey. International Journal of Innovations in Engineering and Technology (IJIET), 1(2), 7-14.

[26] Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry, 27(2), 130.

[27] Féraud, R., & Clérot, F. (2002). A methodology to explain neural network classification. Neural networks, 15(2), 237-246.

[28] Vural, M. S., & Gök, M. (2017). Criminal prediction using Naive Bayes theory. Neural Computing and Applications, 28(9), 2581-2592.

[29] Yang, C. C., Soh, C. S., & Yap, V. V. (2018). A systematic approach in appliance disaggregation using k-nearest neighbours and naive Bayes classifiers for energy efficiency. Energy Efficiency, 11(1), 239-259.

[30] Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion matrix-based feature selection. MAICS, 710, 120-127.

[31] CS6220: DATA MINING TECHNIQUES, Instructor: Yizhou Sun, Chapter 8&9: Classification: Part 3, UCLA Computer Science

[32] Koyejo, O. O., Natarajan, N., Ravikumar, P. K., & Dhillon, I. S. (2014). Consistent binary classification with generalized performance metrics. Advances in neural information processing systems, 27.

[33] Vembandasamy, K., Sasipriya, R., & Deepa, E. (2015). Heart diseases detection using Naive Bayes algorithm. International Journal of Innovative Science, Engineering & Technology, 2(9), 441-444.

[34] Xu, Y., Zhao, X., Chen, Y., & Yang, Z. (2019). Research on a mixed gas classification algorithm based on extreme random tree. Applied Sciences, 9(9), 1728.

[35] Muda, Z., Yassin, W., Sulaiman, M. N., & Udzir, N. I. (2011, December). Intrusion detection based on k-means clustering and OneR classification. In 2011 7th International Conference on Information Assurance and Security (IAS) (pp. 192-197). IEEE.

[36] Muda, Z., Yassin, W., Sulaiman, M. N., & Udzir, N. I. (2011, December). Intrusion detection based on k-means clustering and OneR classification. In 2011 7th International Conference on Information Assurance and Security (IAS) (pp. 192-197). IEEE.

[37] Forcier, J., Bissex, P., & Chun, W. J. (2008). Python web development with Django. Addison-Wesley Professional.

[38] Spurlock, J. (2013). Bootstrap: responsive web development. " O'Reilly Media, Inc.".

[39] Oliphant, T. E. (2006). A guide to NumPy (Vol. 1, p. 85). USA: Trelgol Publishing.

[40] McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. Python for high performance and scientific computing, 14(9), 1-9.

[41] Seabold, S., & Perktold, J. (2010, June). Statsmodels: Econometric and statistical modeling with python. In Proceedings of the 9th Python in Science Conference (Vol. 57, p. 61).

[42] Podrzaj, P. (2019, August). A brief demonstration of some Python GUI libraries. In Proceedings of the 8th International Conference on Informatics and Applications ICIA2019 (pp. 1-6).

[43] Gift, N., & Jones, J. M. (2008). Python for Unix and Linux system administration. " O'Reilly Media, Inc.".

[44] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), 10-18.