

مشروع (فصلي) – قدم استكمالاً لمتطلبات الحصول على درجة البكالوريوس في  
كلية الهندسة – قسم هندسة الذكاء الصناعي وعلوم البيانات

## Predicting energy consumption in buildings

### التنبؤ باستهلاك الطاقة في المباني

إعداد:

راما جمعه

إشراف :

الدكتور: كادان الجمعه

المهندسة: وسام السحلي

العام الدراسي: ٢٠٢٦-٢٠٢٥

شهادة مشرف:

اسم: .....

التاريخ:

التوقيع:

٧	الفصل الأول
٨	١ مقدمة
٨	١,١ خلفية البحث
٨	١,٢ مشكلة البحث
١٠	١,٣ مسائل البحث Research Questions
١١	٤,١ أثر وأهمية البحث Relevance and Importance of the Research
١٢	الفصل الثاني
١٣	٢- الدراسة المرجعية
١٩	الفصل الثالث
٢٠	٣- المعطيات الأولية وطرق تحصيلها
٢٠	٣,١ مجموعة المعطيات Appliances Energy Prediction Dataset
٢١	٢,٣ جدول المتغيرات (Features Description)
٢٣	الفصل الرابع
٢٤	٤- الدراسة النظرية
٢٤	٤,١ التعلّم الآلي (Machine Learning)
٢٥	٤,١,١ أنواع التعلّم الآلي
٢٦	٤,٢ مفهوم الانحدار (Regression)
٢٧	٤,٢,١ مفهوم تحليل الانحدار
٢٧	٤,٢,٢ أنواع الانحدار
٢٩	٤,٣ خوارزميات التعلّم الآلي المستخدمة في الانحدار
٢٩	٤,٣,١ الانحدار الخطي (Linear Regression)
٢٩	٤,٣,٢ الغابات العشوائية (Random Forest Regressor)
٣٠	٤,٣,٣ خوارزمية XGBoost (Extreme Gradient Boosting)
٣٠	٤,٣,٤ انحدار متجهات الدعم (Support Vector Regression – SVR)
٣١	الفصل الخامس
٣٢	٥- المقاربة المقترحة ومنهجية العمل
٣٢	٥,١ مقدمة
٣٢	٥,٢ المعالجة المسبقة للبيانات :
٣٤	٥,٣ هندسة السمات (Feature Engineering)

٣٤	٥,٤ تقسيم البيانات وتطبيقاتها :
٣٥	٥,٥ بناء نماذج التعلم الآلي :
٣٦	٥,٥,١ النماذج المستخدمة في التنبؤ :
٣٦	٥,٦ تقييم النماذج واختيار النموذج الأفضل :
٣٧	الفصل السادس
٣٨	٦-النتائج و تحليلها
٣٨	٦,١ مقدمة :
٣٨	٦,٢ مقاييس تقييم الأداء :
٣٩	٦,٣ نتائج النماذج :
٤٥	٦,٤ تحليل النتائج :
٤٥	٦,٥ مقارنة مع دراسات سابقة :
٤٦	٦,٦ التحديات والتحسينات المحتملة:
٤٧	الفصل السابع
٤٨	٧-الخاتمة و التوصيات
٤٨	٧,١ الخاتمة:
٤٨	٧,٢ التوصيات :
٤٩	٧,٣ افاق مستقبلية :
٥٠	٨-المراجع References

## المصطلحات والتعاريف:

المصطلح التقني	الترجمة العربية	الاختصار	المعنى
<b>Artificial Intelligence</b>	الذكاء الصناعي	AI	مجال من مجالات علوم الحاسب يركز على بناء أنظمة قادرة على أداء مهام تتطلب عادة ذكاءً بشرياً.
<b>Machine Learning</b>	التعلم الآلي	ML	أحد فروع الذكاء الصناعي التي تهتم بتصميم وتطوير خوارزميات وتقنيات تسمح للحواسيب بامتلاك خاصية التعلم.
<b>Deep Learning</b>	التعلم العميق	DL	مجال فرعي من تعلم الآلة يستخدم عدة طبقات مخفية في الشبكات العصبونية لحل المشكلات المعقدة عن طريق تحديد أهم السمات للمعطيات.
<b>Time Series Data</b>	البيانات الزمنية	—	نوع من البيانات يتم جمعه وترتيبه وفق تسلسل زمني، حيث تعتمد القيم الحالية على القيم السابقة.
<b>Long-Short-Term Memory</b>	الذاكرة طويلة وقصيرة الأمد	LSTM	نموذج متقدم من الشبكات العصبية المتكررة قادر على الاحتفاظ بالمعلومات الزمنية المهمة لفترات طويلة والتعامل مع الاعتماديات طويلة المدى..
<b>Feature Engineering</b>	هندسة الميزات	—	عملية استخراج أو إنشاء متغيرات جديدة من البيانات الخام بهدف تحسين أداء نماذج التعلم الآلي.
<b>Lag Features</b>	الميزات المتأخرة	-	ميزات تمثل القيم السابقة للبيانات الزمنية وتُستخدم لمساعدة النموذج على فهم الاعتماديات الزمنية.
<b>Gated Recurrent Unit</b>	وحدة التكرار ذات البوابات	GRU	نموذج من الشبكات العصبية المتكررة يُستخدم لمعالجة البيانات الزمنية، ويتميز ببنية أبسط من LSTM مع قدرة فعالة على تمثيل الاعتماديات الزمنية
<b>Extreme-Gradient Boosting</b>	التعزيز التدرجي المتطرف	XGBoost	خوارزمية تعلم آلي تعتمد على تجميع الأشجار القرارية بطريقة معززة، وتتميز بالكفاءة العالية والدقة في مسائل التنبؤ.
<b>Linear Regression</b>	الانحدار الخطي	—	نموذج إحصائي من نماذج تعلم الآلة يُستخدم لتقدير العلاقة الخطية بين المتغير الهدف ومجموعة من المتغيرات المستقلة.
<b>Random Forest Regressor</b>	الغابة العشوائية للانحدار	—	نموذج تعلم آلي يعتمد على إنشاء مجموعة من أشجار القرار ودمج نتائجها لتحسين دقة التنبؤ وتقليل التحيز والتباين.

مقياس الفرق بين القيم المتوقعة والقيم الفعلية.	RMSE	جذر متوسط الخطأ التربيعي	<b>Root Mean Square Error</b>
مقياس إحصائي يُستخدم لتقييم أداء نماذج الانحدار، ويعبر عن نسبة التباين في المتغير الهدف التي يفسرها النموذج.	$R^2$	معامل التحديد	<b>Coefficient of Determination</b>
خوارزمية لتصنيف المعطيات.	SVM	آلة المتجه الداعم	<b>Support Vector Machine</b>

# الفصل الأول

# ١ مقدمة

## ١,١ خلفية البحث

في ظل التحوّلات الرقمية السريعة، برزت تقنيات الذكاء الصناعي (Artificial Intelligence) وتعلّم الآلة (Machine Learning) كعوامل رئيسية تسهم في تحسين إدارة الموارد والطاقة، ولا سيّما في قطاع المباني الذي يُعد من أكثر القطاعات استهلاكاً للطاقة الكهربائية.

وقد أدى التوسع في جمع البيانات وتوفّرها إلى إتاحة فرص جديدة لتحليل أنماط استهلاك الطاقة والتنبؤ بها، مما يساهم في تحسين كفاءة الاستخدام، تقليل الهدر، ودعم التوجّه نحو أنظمة طاقة أكثر استدامة.

ويُعد التنبؤ باستهلاك الطاقة في المباني من التطبيقات المهمة التي توفّر حلولاً عملية لمواجهة التحديات المرتبطة بارتفاع الطلب على الطاقة وتزايد التكاليف التشغيلية.

ومن أبرز التقنيات التي تدعم تطوير أنظمة التنبؤ الحديثة هي تقنيات تعلّم الآلة وتحليل السلاسل الزمنية، والتي أحدثت نقلة نوعية في قدرة الأنظمة على معالجة البيانات التاريخية واستخلاص الأنماط الزمنية المعقّدة.

فقد أتاح استخدام نماذج تعلّم الآلة، مثل نماذج الانحدار والخوارزميات المعزّزة والغابات العشوائية، إمكانية تمثيل العلاقات غير الخطية بين العوامل الزمنية والبيئية المؤثرة في استهلاك الطاقة.

كما ساهمت تقنيات هندسة الميزات الزمنية في تحسين قدرة النماذج على فهم الدورات اليومية والموسمية لأنماط الاستهلاك، مما يتيح بناء أنظمة تنبؤية أكثر دقة وموثوقية تدعم قرارات إدارة الطاقة في المباني بشكل فعّال.

## ١,٢ مشكلة البحث

على الرغم من التطورات الكبيرة في مجال إدارة الطاقة الذكية واستخدام تقنيات الذكاء الصناعي في تحليل البيانات، فإن التنبؤ الدقيق باستهلاك الطاقة في المباني لا يزال يواجه العديد من التحديات.

ويعود ذلك إلى الطبيعة المعقّدة لاستهلاك الطاقة، حيث يتأثر بعوامل متعددة ومتداخلة تشمل العوامل الزمنية، السلوكية، والبيئية، مما يزيد من صعوبة تطوير نماذج تنبؤية دقيقة وموثوقة.



ففي حين تحقق بعض أنظمة التنبؤ أداءً جيداً في بيئات محددة، إلا أن تعميم هذه النماذج على بيانات زمنية حقيقية طويلة المدى لا يزال محدوداً بسبب التقلبات وعدم الاستقرار في أنماط الاستهلاك.

ومن أهم التحديات التي تواجه التنبؤ باستهلاك الطاقة هو تعقيد البيانات الزمنية نفسها، حيث تتسم بوجود أنماط دورية يومية وموسمية، إضافةً إلى تقلبات مفاجئة ناتجة عن سلوك المستخدمين أو الظروف المناخية المتغيرة.

يؤثر هذا التعقيد بشكل مباشر على قدرة النماذج التنبؤية على تمثيل العلاقات بدقة، مما يجعل من الصعب تحقيق أداء ثابت وموثوق عبر فترات زمنية مختلفة.

وقد تم توظيف عدة مقاربات لمعالجة هذه المشكلة، من بينها:

### ١. نماذج تعلم الآلة التقليدية: (Machine Learning)

تعتمد هذه المقاربة على خوارزميات تعلم الآلة مثل الانحدار الخطي، الغابات العشوائية، ونماذج التعزيز، حيث يتم تدريب النماذج على بيانات تاريخية لاستخلاص الأنماط المؤثرة في استهلاك الطاقة.

وعلى الرغم من بساطة هذه النماذج وكفاءتها في بعض الحالات، إلا أنها تعاني من محدودية في التعامل مع العلاقات غير الخطية والاعتماديات الزمنية المعقدة.

### ٢. نماذج التعزيز المعتمدة على الأشجار: (Ensemble Learning)

تعتمد هذه المقاربة على دمج عدة نماذج تنبؤية، مثل خوارزميات التعزيز التدريجي، بهدف تحسين دقة التنبؤ وتقليل التحيز والتباين.

وقد أثبتت هذه النماذج فعاليتها في التعامل مع البيانات غير الخطية، إلا أن أدائها يعتمد بشكل كبير على جودة الميزات المستخرجة من البيانات الزمنية.

### ٣. هندسة الميزات الزمنية: (Time Series Feature Engineering)

تشمل هذه التقنيات استخراج ميزات مشتقة من الزمن مثل القيم المتأخرة (Lag Features)، المتوسطات المتحركة، والترميز الدوري للمتغيرات الزمنية.

تسهم هذه المقاربة في تحسين قدرة النماذج على فهم البنية الزمنية للبيانات، غير أن اختيار الميزات المناسبة يبقى تحدياً أساسياً يؤثر بشكل مباشر على أداء النموذج.

#### ٤. تقسيم البيانات الزمني: (Time-Based Data Splitting)

تُعد هذه المقاربة ضرورية لضمان تقييم واقعي لأداء نماذج التنبؤ، حيث يتم تقسيم البيانات اعتماداً على التسلسل الزمني بدلاً من التقسيم العشوائي.

ورغم أهميتها، فإنها تفرض قيوداً إضافية على عملية التدريب، وتزيد من صعوبة تحقيق أداء مرتفع عند التنبؤ بالقيم المستقبلية.

#### ٥. تقييم النماذج باستخدام مقاييس إحصائية مناسبة:

يتطلب التنبؤ باستهلاك الطاقة استخدام مقاييس تقييم دقيقة مثل متوسط الخطأ المطلق (MAE)، والجذر التربيعي لمتوسط مربع الخطأ (RMSE)، ومعامل التحديد ( $R^2$ )، وذلك لضمان تقييم شامل يعكس دقة النموذج وقدرته على تفسير التباين في بيانات الاستهلاك.

بناءً على التحديات التي تم توضيحها، تبرز الحاجة إلى تطوير نماذج تنبؤية قائمة على تقنيات تعلم الآلة قادرة على التعامل مع الطبيعة الزمنية المعقدة لاستهلاك الطاقة في المباني، بما يسهم في تحسين دقة التنبؤ ودعم أنظمة إدارة الطاقة الذكية بشكل أكثر فعالية.

### ١,٣ مسائل البحث Research Questions

بناءً على مشكلة البحث المتعلقة بالتنبؤ باستهلاك الطاقة في المباني باستخدام تقنيات تعلم الآلة، وأهمية تمثيل البيانات الزمنية والبيئية بشكل فعال لتعزيز دقة وموثوقية التنبؤ، يمكن صياغة المسائل البحثية النظرية والعملية المطلوبة كما يلي:

١. هل يمكن لنماذج تعلم الآلة التقليدية التنبؤ باستهلاك الطاقة في المباني بدقة كافية اعتماداً على البيانات الزمنية والبيئية المتاحة؟

٢. دراسة تأثير هندسة الميزات الزمنية، مثل القيم المتأخرة والمتوسطات المتحركة والترميز الدوري للزمن، على أداء نماذج التنبؤ باستهلاك الطاقة.

٣. دراسة النماذج المستخدمة حالياً في التنبؤ باستهلاك الطاقة، ومقارنة أدائها بهدف تحديد النموذج الأكثر كفاءة ودقة في تمثيل أنماط الاستهلاك في المباني.

## ٤,١ أثر وأهمية البحث Relevance and Importance of the Research

يهدف هذا البحث إلى تقديم إسهامات نوعية في مجال التنبؤ باستهلاك الطاقة في المباني، وذلك من خلال تطوير نماذج تعتمد على تقنيات تعلم الآلة وتحليل البيانات الزمنية، مع دمج سمات مستخرجة من البيانات البيئية والزمنية تهدف إلى تعزيز قدرة النموذج على تمثيل أنماط الاستهلاك والتغيرات الدورية في الطلب على الطاقة.

يعكس البحث أهمية علمية وتطبيقية في آنٍ معاً، حيث يساهم في إثراء المعرفة الأكاديمية في مجال نمذجة استهلاك الطاقة، ويعالج التحديات المرتبطة بتقلب الأحمال الطاقية وتعقيد العوامل المؤثرة عليها في المباني.

على المستوى النظري، يساهم البحث في دفع عجلة التطوير العلمي في مجال تحليل السلاسل الزمنية وتطبيقات تعلم الآلة في أنظمة الطاقة، والذي لا يزال يواجه تحديات تتعلق بدقة التنبؤ واستقرار النماذج في البيئات الواقعية.

ومن خلال هذا البحث، يُتوقع استكشاف أثر هندسة الميزات الزمنية، مثل القيم المتأخرة، المتوسطات المتحركة، والترميز الدوري للزمن، على تحسين أداء نماذج التنبؤ.

كما يتيح ذلك دراسة مقارنة بين نماذج مختلفة من تعلم الآلة، مما يساهم في بناء نموذج تنبؤي أكثر قدرة على تمثيل أنماط الاستهلاك اليومية والموسمية للطاقة.

أما على المستوى التطبيقي، فإن البحث يحمل أهمية كبيرة لقطاعات إدارة الطاقة والمباني الذكية، التي تتطلب أدوات دقيقة وموثوقة للتنبؤ باستهلاك الطاقة بهدف تحسين كفاءة التشغيل وتقليل الهدر.

يساهم النموذج المقترح في دعم قرارات إدارة الأحمال الكهربائية، وتخطيط استهلاك الطاقة، وتعزيز استراتيجيات الترشيد، مما يؤدي إلى خفض التكاليف التشغيلية وتحسين الاستدامة البيئية.

إضافة إلى ذلك، يمكن الاستفادة من نتائج هذا البحث في تطوير أنظمة إدارة الطاقة الذكية التي تتكامل مع أنظمة المراقبة والتحكم في المباني الحديثة.

بإيجاز، يسعى هذا البحث إلى سد فجوة معرفية وتطبيقية في مجال التنبؤ باستهلاك الطاقة في المباني، من خلال تقديم نموذج تنبؤي قائم على تعلم الآلة يمكن أن يشكل مرجعاً بحثياً وتطبيقياً لتطوير حلول ذكية تدعم كفاءة استخدام الطاقة والاستدامة في البيئات العمرانية الحديثة.

## الفصل الثاني

## ٢- الدراسة المرجعية

### ٢,١ المراجعات السابقة

شهدت السنوات الأخيرة تطوراً ملحوظاً في توظيف تقنيات التعلم الآلي في مجال التنبؤ باستهلاك الطاقة في المباني، مدفوعاً بالحاجة المتزايدة إلى تحسين كفاءة إدارة الطاقة وتقليل الهدر ودعم مفاهيم الاستدامة البيئية.

وقد ساهم التقدم في قدرات المعالجة الحاسوبية وتوفر البيانات الزمنية والبيئية عالية الدقة في فتح آفاق جديدة لتطبيق خوارزميات تعلم الآلة في تحليل أنماط الاستهلاك الطاقوي وبناء نماذج تنبؤية أكثر دقة وموثوقية.

وأظهرت الدراسات الحديثة أن نماذج التعلم الآلي، بما في ذلك نماذج الانحدار، والنماذج التجميعية مثل Random Forest و XGBoost، قادرة على تحقيق أداء متقدم مقارنة بالأساليب الإحصائية التقليدية، ولا سيما في التعامل مع العلاقات غير الخطية والتقلبات الزمنية المعقدة.

كما بينت بعض الأبحاث أن النماذج المتقدمة المعتمدة على التعلم العميق، مثل LSTM والنماذج الهجينة، تمثل امتداداً طبيعياً لتقنيات التعلم الآلي عند معالجة السلاسل الزمنية طويلة الأمد والأنظمة ذات السلوك الديناميكي.

ورغم التطور الكبير الذي حققته نماذج التعلم الآلي في مجال التنبؤ باستهلاك الطاقة في المباني، فقد أشارت الدراسات السابقة إلى أن أداء هذه النماذج يرتبط ارتباطاً وثيقاً بطبيعة البيانات الزمنية المستخدمة.

إذ بينت الأبحاث أن بيانات استهلاك الطاقة تنتم بوجود أنماط دورية واعتماديات زمنية بين القيم المتتالية، مما يجعل عملية التنبؤ أكثر تعقيداً مقارنة بالبيانات التقليدية، ويستدعي استخدام نماذج قادرة على التعامل مع هذا النوع من البيانات بكفاءة.

ومن هذا المنطلق، تهدف الدراسات المرجعية في هذا الفصل إلى استعراض وتحليل أبرز الأبحاث التي تناولت التنبؤ باستهلاك الطاقة في المباني باستخدام تقنيات التعلم الآلي، مع التركيز على النماذج المستخدمة، أنواع البيانات، وأساليب التقييم المعتمدة.

ويسهم هذا الاستعراض في إبراز الاتجاهات البحثية الحديثة، وتحديد الفجوات العلمية القائمة، وتوفير أرضية معرفية متينة لدعم المقاربة المقترحة في هذا البحث.

## ٢,٢ الدراسة المرجعية :

أبرزت الدراسات الحديثة الأهمية المتزايدة لاستخدام تقنيات التعلم الآلي (Machine Learning) في مجال التنبؤ باستهلاك الطاقة في المباني، نظراً لقدرتها على تحسين كفاءة إدارة الطاقة، تقليل الهدر، ودعم التوجّه نحو أنظمة طاقة أكثر استدامة.

وقد أظهرت هذه الدراسات فعالية خوارزميات التعلم الآلي في التعامل مع البيانات الزمنية والمعيارية ذات الطبيعة المعقّدة، من خلال نمذجة العلاقات الخطية وغير الخطية بين المتغيرات المؤثرة في استهلاك الطاقة.

كما تتوّعت الأساليب المستخدمة بين نماذج الانحدار التقليدية، والنماذج التجميعية (Ensemble Models) مثل Random Forest و XGBoost ، إضافة إلى بعض النماذج المتقدمة المعتمدة على التعلم العميق بوصفها امتداداً لتقنيات التعلم الآلي في معالجة الاعتماديات الزمنية المعقّدة.

ركّزت الدراسة الأولى [١] على استخدام نماذج تعلم الآلة التقليدية لتحليل استهلاك الطاقة في منزل ، بعد معالجة البيانات البيئية والسلوكية للأجهزة والإضاءة و درجات الحرارة و الرطوبة، إضافة إلى بيانات الطقس الخارجي.

اعتمدت الدراسة على بيانات Appliances Energy Prediction Dataset، والتي أتيح من خلالها استخراج المتغيرات الأكثر تأثيراً على الاستهلاك.

وأجريت الدراسة اختبار أربعة نماذج تنبؤية ضمن إطار تعلم الآلة :الانحدار الخطي المتعدد (LM) ، وآلة المتجهات الداعمة ذات النواة الشعاعية (SVM Radial) ، وآلات تعزيز التدرج (GBM) ، والغابة العشوائية.(RF)

وأظهرت النتائج أن النماذج التجميعية، خصوصاً GBM و RF، قدمت أداءً أفضل مقارنة بالانحدار الخطي و SVM، حيث سجل نموذج GBM في مجموعة التدريب  $MAPE = 16.27\%$  ،  $MAE = 11.97$  ،  $R^2 = 0.97$  ،  $RMSE = 17.56$  ، وفي مجموعة الاختبار  $MAPE = 38.29\%$  ،  $MAE = 35.22$  ،  $R^2 = 0.57$  ،  $RMSE = 66.65$  ، بينما سجلت RF في مجموعة الاختبار  $MAPE = 31.39\%$  ،  $MAE = 31.85$  ،  $R^2 = 0.54$  ،  $RMSE = 68.48$

وتؤكد هذه النتائج أن خوارزميات تعلم الآلة التجميعية قادرة على تمثيل العلاقات المعقدة بين المتغيرات التشغيلية والبيئية بشكل أفضل من النماذج التقليدية، مما يجعلها أدوات فعالة للتنبؤ الدقيق باستهلاك الطاقة في البيئات المنزلية.

بينما ركّزت الدراسة الثانية [٢] على استخدام نموذج LSTM ضمن إطار التعلم الآلي لتحليل البيانات الزمنية بعد إجراء معالجة متقدمة للبيانات باستخدام تقنيات هندسة الميزات (Feature Engineering) وتحليل المكونات الرئيسية.(PCA)

اعتمدت الدراسة على بيانات NREL Research Support Facility (RSF) في الولايات المتحدة، والتي شملت متغيرات تشغيلية متعددة مثل استهلاك الأجهزة، الإضاءة، ودرجات الحرارة.

وأظهرت النتائج أن المعالجة الدقيقة للبيانات أسهمت في تحسين أداء النموذج بشكل ملحوظ، حيث حقق  $R^2 = 0.97$  و  $RMSE = 0.02$  و  $MAPE = 0.03$ ، مما يؤكد أن نماذج LSTM تمثل امتداداً متقدماً لتقنيات التعلم الآلي التقليدية عند التعامل مع العلاقات الزمنية المعقدة.

أما الدراسة الثالثة [3] فقد تناولت التنبؤ قصير الأمد لاستهلاك الطاقة في أنظمة التدفئة والتبريد (HVAC) باستخدام نموذج Attention-Based LSTM، الذي يُعد تطويراً لنماذج التعلم الآلي الزمنية من خلال دمج آلية الانتباه لتحسين دقة التنبؤ في الفترات التي تشهد تغيرات حادة في الأحمال الطاقة.

استخدمت الدراسة بيانات نظام CCHP لمبنى جامعي في اليابان، وأظهرت النتائج تفوق النموذج المقترح على الأساليب التقليدية، حيث حقق  $RMSE = 74.97 \text{ kW}$  و  $MAE = 44.79 \text{ kW}$  و معامل تحديد  $R^2 = 0.929$ ، مما يعكس قدرة تقنيات التعلم الآلي المتقدمة على التكيف مع السلوك الديناميكي المعقد لاستهلاك الطاقة.

وتوسّعت الدراسة الرابعة [4] لتناول مشكلة التنبؤ باستهلاك الطاقة على نطاق وطني واسع، من خلال تطبيق نموذج هجين يجمع بين الشبكة الالتفافية CNN وشبكة الذاكرة الطويلة القصيرة LSTM، اعتماداً على بيانات متعددة المصادر في كوريا الجنوبية خلال الفترة الممتدة بين عامي ٢٠٠٤ و ٢٠٢٠.

شملت البيانات استهلاك الكهرباء، بيانات الطقس، ومؤشرات التجارة.

أظهرت النتائج تفوق النموذج الهجين على النماذج التقليدية مثل MLP و RNN و LSTM، حيث حقق

$RMSE = 0.1650$  و  $MAPE = 0.1441$ ، مما يبيّن أن دمج تقنيات التعلم الآلي التقليدية مع نماذج أكثر تقدماً يسهم في تحسين دقة التنبؤ عبر التقاط الأنماط الزمنية والمكانية المتداخلة.

كما ركّزت الدراسة الخامسة [5] على دمج تقنيات إنترنت الأشياء (IoT) مع خوارزميات التعلم الآلي لتعزيز دقة التنبؤ باستهلاك الطاقة في المباني الذكية، وذلك من خلال نموذج هجين يعتمد على CNN-LSTM ضمن بنية تكاملية مدعومة ببطاقة تكامل IoT.

استخدمت الدراسة بيانات حقيقية من مستشعرات في مبانٍ سكنية وتجارية، وأظهرت النتائج تفوق النموذج الهجين على النماذج الفردية، حيث حقق  $RMSE = 1.72$  و  $MAE = 1.45$  ومعامل تحديد مرتفع  $R^2 = 0.983$ ، مما يؤكد فعالية توظيف تقنيات التعلم الآلي في البيئات الذكية المعتمدة على البيانات الحية.

وقدّمت الدراسة السادسة [٦] مراجعة شاملة لأبحاث التعلّم الآلي المعتمدة على البيانات (Data-Driven Frameworks) في مجال التنبؤ باستهلاك الطاقة، حيث استعرضت أكثر من ١٩٠ دراسة بين عامي ٢٠٠٠ و ٢٠٢٥، شملت نماذج تقليدية مثل Linear Regression و SVR و ARIMA، إلى جانب النماذج التجميعية والهجينة مثل XGBoost و CNN – LSTM. وأكدت الدراسة أن خوارزميات التعلّم الآلي، ولا سيّما النماذج التجميعية والهجينة، تحقق أعلى مستويات الدقة في التنبؤ قصير الأمد، كما أشارت إلى أن دمج تقنيات IoT والتعلّم النقلي (Transfer Learning) والذكاء القابل للتفسير (Explainable AI) يمثل اتجاهاً بحثياً واعداً في هذا المجال.

وبشكل عام، أجمعت الدراسات الست على أن تقنيات التعلّم الآلي تمثل الركيزة الأساسية للتنبؤ الدقيق باستهلاك الطاقة في المباني، مع إمكانية الاستفادة من النماذج الهجينة والتعلّم العميق كخيارات متقدمة لتحسين الأداء في الأنظمة المعقّدة. كما أثبتت النتائج أن استخدام النماذج التجميعية وهندسة الميزات الزمنية يوفر أدوات فعّالة لتصميم أنظمة إدارة طاقة ذكية تدعم الكفاءة التشغيلية والاستدامة البيئية.

النتائج	معايير التقييم	النماذج والتقنيات	DATASET المستخدمة	الدراسة
1. نموذج الانحدار الخطي المتعدد (LM): RMSE: 93.18 - R2: 0.16 - MAE: 51.97 - MAPE %: 59.93 - 2. نموذج آلة المتجهات الداعمة ذات النواة الشعاعية (SVM Radial): RMSE: 70.74 - R2: 0.52 - MAE: 31.36 - MAPE %: 29.76 - 3. نموذج آلات تعزيز التدرج (GBM): RMSE: 66.65 - R2: 0.57 - MAE: 35.22 - MAPE %: 38.29 - 4. نموذج الغابة العشوائية (RF): - مجموعة التدريب: RMSE: 29.61 - R2: 0.92 - MAE: 13.75 -	<ul style="list-style-type: none"> <li>• RMSE (Root Mean Square Error)</li> <li>• MAE (Mean Absolute Error)</li> <li>• R<sup>2</sup> (Coefficient of Determination)</li> </ul>	<ul style="list-style-type: none"> <li>• Multiple Linear Regression (MLR)</li> <li>• Support Vector Machine (SVM)</li> <li>• Random Forest (RF)</li> <li>• Gradient Boosting Machine (GBM)</li> </ul>	Appliances Energy Prediction Dataset	<p><u><a href="#">“Data-driven prediction models of energy use of appliances in a low-energy house”</a></u></p>



				MAPE %: 13.43 - مجموعة الاختبار: RMSE: 68.48 - R2: 0.54 - MAE: 31.85 - MAPE %: 31.39 -
<a href="#">An Analysis of the Energy Consumption Forecasting Problem in Smart Buildings Using LSTM</a>	Raw_Data.csv Research Support Facility (RSF) U.S. National Renewable Energy Laboratory (NREL)	<ul style="list-style-type: none"> <li>Long Short-Term Memory (LSTM)</li> <li>Feature Engineering (تحليل المتغيرات عبر: Pearson &amp; Spearman Correlation, Multiple Linear Regression, PCA, ARIMA, Autocorrelation)</li> <li>Principal Component Analysis (PCA) لاستخلاص المكونات الرئيسية (PC1-PC6)</li> <li>ARIMA Models لتحديد التأخر الزمني (lags)</li> <li>Normalization (Min-Max)</li> </ul>	<ul style="list-style-type: none"> <li>RMSE (Root Mean Square Error)</li> <li>MAPE (Mean Absolute Percentage Error)</li> <li>R<sup>2</sup> (Coefficient of Determination)</li> </ul>	المجموعة 1 (PC1 + Skyspark): RMSE = 0.07, MAPE = 0.10, R <sup>2</sup> = 0.74 المجموعة 2 (PC1 + PC2 + Skyspark): RMSE = 0.04, MAPE = 0.06, R <sup>2</sup> = 0.88 المجموعة 3 RMSE = 0.02, MAPE = 0.03, R <sup>2</sup> = 0.97
<a href="#">Potential Analysis of the Attention-Based LSTM Model in Ultra-Short-Term Forecasting of Building HVAC Energy Consumption</a>	Kitakyushu Science and Research Park (KSRP) – اليابان بيانات نظام CCHP (Combined Cooling, Heating, and Power) في مبنى جامعي	<ul style="list-style-type: none"> <li>Attention-Based LSTM (A-LSTM)</li> <li>Long Short-Term Memory (LSTM)</li> <li>Recurrent Neural Network (RNN)</li> <li>Deep Neural Network (DNN)</li> <li>Support Vector Regression (SVR)</li> </ul>	<ul style="list-style-type: none"> <li>RMSE (Root Mean Square Error)</li> <li>MAE (Mean Absolute Error)</li> <li>R<sup>2</sup> (Coefficient of Determination)</li> </ul>	الأداء السنوي (2010): <ul style="list-style-type: none"> <li>RMSE = 74.977 kW</li> <li>MAE = 44.793 kW</li> <li>R<sup>2</sup> = 0.929</li> </ul> تحسن عن LSTM التقليدي: <ul style="list-style-type: none"> <li>RMSE ↓ 3.06%</li> <li>MSE ↓ 6.54%</li> <li>R<sup>2</sup> ↑ 0.43%</li> </ul>
<a href="#">Accurate prediction of electricity consumption</a>	بيانات حكومية لكوريا الجنوبية	<ul style="list-style-type: none"> <li>MLP – Multi-Layer Perceptron</li> <li>RNN – Recurrent Neural Network</li> <li>LSTM – Long Short-Term</li> </ul>	RMSE (Root Mean Squared Error). - MAPE (Mean Absolute Percentage Error)	<ul style="list-style-type: none"> <li>MLP: RMSE = 0.4521, MAPE = 0.4727</li> <li>RNN: RMSE = 0.1798, MAPE = 0.1602</li> </ul>

<a href="#"><u>using a hybrid CNN-LSTM model based on multivariable data.</u></a>		<ul style="list-style-type: none"> <li>Memory</li> <li>• CNN-LSTM</li> </ul>		<ul style="list-style-type: none"> <li>• LSTM: RMSE = 0.1740 , MAPE = 0.1589</li> <li>• CNN-LSTM: RMSE = 0.1650 , MAPE = 0.1441</li> </ul>
<a href="#"><u>Enhancing Building Energy Efficiency with IoT-Driven Hybrid Deep Learning Models for Accurate Energy Consumption Prediction</u></a>	تم جمعها عبر أجهزة إنترنت الأشياء (IoT Sensors) من مبانٍ ذكية في بيئة واقعية	<ul style="list-style-type: none"> <li>• CNN-Convolutional Neural Network</li> <li>• LSTM – Long Short-Term Memory</li> <li>• CNN-LSTM</li> <li>• IoT Integration Layer</li> <li>• Adam Optimizer, ReLU Activation</li> </ul>	<ul style="list-style-type: none"> <li>• RMSE – Root Mean Square Error</li> <li>• MAE – Mean Absolute Error</li> <li>• <math>R^2</math> – Coefficient of Determination</li> </ul>	<ul style="list-style-type: none"> <li>• LSTM: RMSE = 2.13 , MAE = 1.81 , <math>R^2 = 0.96</math></li> <li>• CNN-LSTM: RMSE = 1.72 , MAE = 1.45 , <math>R^2 = 0.983</math></li> </ul>
<a href="#"><u>A Systematic Review of Building Energy Consumption Prediction: From Perspectives of Load Classification, Data-Driven Frameworks, and Future Directions</u></a>	لم تستخدم الورقة مجموعة بيانات محددة، لأنها دراسة مراجعة شاملة	<ul style="list-style-type: none"> <li>Support Vector Regression (SVR)</li> <li>• Artificial Neural Network (ANN)</li> <li>• Convolutional Neural Network (CNN)</li> <li>• Long Short-Term Memory (LSTM)</li> <li>• Random Forest (RF)</li> <li>• ARIMA (Autoregressive Integrated Moving Average)</li> <li>• Hybrid &amp; Transfer Learning Models</li> </ul>	<ul style="list-style-type: none"> <li>• RMSE (Root Mean Square Error)</li> <li>• MAE (Mean Absolute Error)</li> <li>• MAPE (Mean Absolute Percentage Error)</li> <li>• <math>R^2</math> (Coefficient of Determination)</li> </ul>	<ul style="list-style-type: none"> <li>• LSTM و CNN أكثر النماذج دقة في التنبؤ الزمني لاستهلاك الطاقة.</li> <li>• الجمع بين النماذج (Hybrid Models) يعطي أداءً أفضل من النماذج المنفردة.</li> </ul>

## الفصل الثالث

## ٣-المعطيات الأولية وطرق تحصيلها

### ٣,١ مجموعة المعطيات Appliances Energy Prediction Dataset :

المعطيات المستخدمة في هذا البحث هي مجموعة معطيات **Appliances Energy Prediction**، وهي مجموعة معطيات حقيقية تم جمعها من منزل سكني يقع في بلجيكا، وتهدف إلى تحليل أنماط استهلاك الطاقة الكهربائية للأجهزة المنزلية والتنبؤ بها اعتماداً على مجموعة من العوامل الزمنية والبيئية.

تم جمع هذه المعطيات ضمن إطار دراسة علمية سابقة، وهي متاحة للاستخدام البحثي في مجال إدارة الطاقة والمباني الذكية.

تتضمن مجموعة المعطيات قياسات لاستهلاك الطاقة الكهربائية للأجهزة المنزلية بوحدة الواط-ساعة (Wh) ، حيث تم تسجيل القراءات كل ١٠ دقائق على مدى فترة زمنية طويلة، مما يوفر سلسلة زمنية عالية الدقة تسمح بتحليل التغيرات الدقيقة في سلوك استهلاك الطاقة.

إضافة إلى ذلك، تحتوي المعطيات على مجموعة من المتغيرات البيئية والفيزيائية التي تم جمعها بواسطة حساسات موزعة داخل المنزل وخارجه.

تشمل هذه المتغيرات درجات الحرارة والرطوبة في عدة غرف داخل المبنى، مثل المطبخ وغرف المعيشة، إلى جانب درجة الحرارة والرطوبة الخارجية، الأمر الذي يتيح دراسة العلاقة بين الظروف البيئية واستهلاك الطاقة الكهربائية للأجهزة المنزلية. ويساهم هذا التنوع في المتغيرات في تحسين قدرة نماذج تعلم الآلة على فهم الأنماط المعقدة المرتبطة باستهلاك الطاقة في البيئات السكنية.

يمثل الاعتماد على بيانات حقيقية مأخوذة من منزل سكني فعلي عنصراً أساسياً في موثوقية هذا البحث، حيث تعكس البيانات سلوكاً واقعياً غير منتظم ناتجاً عن الاستخدام البشري اليومي للأجهزة الكهربائية.

كما أن تسجيل القراءات بشكل آلي ومنتظم كل ١٠ دقائق يقلل من احتمالية الأخطاء البشرية ويعزز من دقة القياسات.

تُعد هذه المجموعة من المعطيات مناسبة بشكل خاص لتطبيق نماذج التنبؤ المعتمدة على السلاسل الزمنية، إذ تتيح دراسة الأنماط قصيرة الأمد (داخل اليوم الواحد) والأنماط طويلة الأمد (اليومية والموسمية). كما تُمكن من تطبيق تقنيات متقدمة في

هندسة الميزات الزمنية، مثل القيم المتأخرة والمتوسطات المتحركة والترميز الدوري للزمن، مما يدعم تطوير نماذج تنبؤية أكثر دقة وفعالية في مجال التنبؤ باستهلاك الطاقة في المباني.

## ٢.٣ جدول المتغيرات (Features Description)

اسم المتغير	نوع المتغير	الوصف
Date	زمني (Datetime)	تاريخ ووقت تسجيل القراءة، حيث يتم قياس القيم كل ١٠ دقائق.
Appliances	رقمي مستمر	استهلاك الطاقة الكهربائية للأجهزة المنزلية بوحدة واط-ساعة (Wh)، وهو المتغير الهدف في هذا البحث.
lights	رقمي مستمر	استهلاك الطاقة الخاص بالإضاءة داخل المنزل بوحدة واط-ساعة (Wh).
T1	رقمي مستمر	درجة الحرارة في المطبخ. (°C)
RH_1	رقمي مستمر	الرطوبة النسبية في المطبخ. (%)
T2	رقمي مستمر	درجة الحرارة في غرفة المعيشة. (°C)
RH_2	رقمي مستمر	الرطوبة النسبية في غرفة المعيشة. (%)
T3	رقمي مستمر	درجة الحرارة في غرفة الغسيل. (°C)
RH_3	رقمي مستمر	الرطوبة النسبية في غرفة الغسيل. (%)
T4	رقمي مستمر	درجة الحرارة في المكتب. (°C)
RH_4	رقمي مستمر	الرطوبة النسبية في المكتب. (%)
T5	رقمي مستمر	درجة الحرارة في الحمام. (°C)
RH_5	رقمي مستمر	الرطوبة النسبية في الحمام. (%)
T6	رقمي مستمر	درجة الحرارة خارج المبنى. (°C)
RH_6	رقمي مستمر	الرطوبة النسبية خارج المبنى. (%)
T7	رقمي مستمر	درجة الحرارة في غرفة التخزين. (°C)
RH_7	رقمي مستمر	الرطوبة النسبية في غرفة التخزين. (%)
T8	رقمي مستمر	درجة الحرارة في غرفة الضيوف. (°C)
RH_8	رقمي مستمر	الرطوبة النسبية في غرفة الضيوف. (%)
T9	رقمي مستمر	درجة الحرارة في غرفة النوم. (°C)
RH_9	رقمي مستمر	الرطوبة النسبية في غرفة النوم. (%)
T_out	رقمي مستمر	درجة الحرارة الخارجية المسجلة من محطة الطقس. (°C)
Press_mm_hg	رقمي مستمر	الضغط الجوي بوحدة ميليمتر زئبق (mmHg)
RH_out	رقمي مستمر	الرطوبة النسبية الخارجية. (%)
Windspeed	رقمي مستمر	سرعة الرياح. (m/s)

مدى الرؤية الجوية.(km)	رقمي مستمر	Visibility
درجة حرارة نقطة الندى.(°C)	رقمي مستمر	Tdewpoint

display(df\_clean.head(5))

date	Appliances	lights	T1	RH_1	T2	RH_2	T3	RH_3	T4	RH_4	...	T8	RH_8	T9	RH_9	T_out	Press_mm_hg	RH_out	Windspeed	Visibility	Tdewpoint
2016-01-11 17:00:00	60	30	19.89	47.596667	19.2	44.790000	19.79	44.730000	19.000000	45.566667	...	18.2	48.900000	17.033333	45.53	6.600000	733.5	92.0	7.000000	63.000000	5.3
2016-01-11 17:10:00	60	30	19.89	46.693333	19.2	44.722500	19.79	44.790000	19.000000	45.992500	...	18.2	48.863333	17.066667	45.56	6.483333	733.6	92.0	6.666667	59.166667	5.2
2016-01-11 17:20:00	50	30	19.89	46.300000	19.2	44.626667	19.79	44.933333	18.926667	45.890000	...	18.2	48.730000	17.000000	45.50	6.366667	733.7	92.0	6.333333	55.333333	5.1
2016-01-11 17:30:00	50	40	19.89	46.066667	19.2	44.590000	19.79	45.000000	18.890000	45.723333	...	18.1	48.590000	17.000000	45.40	6.250000	733.8	92.0	6.000000	51.500000	5.0
2016-01-11 17:40:00	60	40	19.89	46.333333	19.2	44.530000	19.79	45.000000	18.890000	45.530000	...	18.1	48.590000	17.000000	45.40	6.133333	733.9	92.0	5.666667	47.666667	4.9

5 rows x 26 columns

الشكل (١)

يوضح الشكل (١) عينة من المعطيات المستخدمة في هذا العمل، حيث تُظهر أول عدد من السجلات مع المتغيرات المرتبطة بها. وتساعد هذه العينة على إعطاء تصور واضح عن بنية مجموعة البيانات، ونوع القيم، وطبيعة المتغيرات الزمنية والبيئية المعتمدة في عملية التنبؤ.

## الفصل الرابع

## ٤- الدراسة النظرية

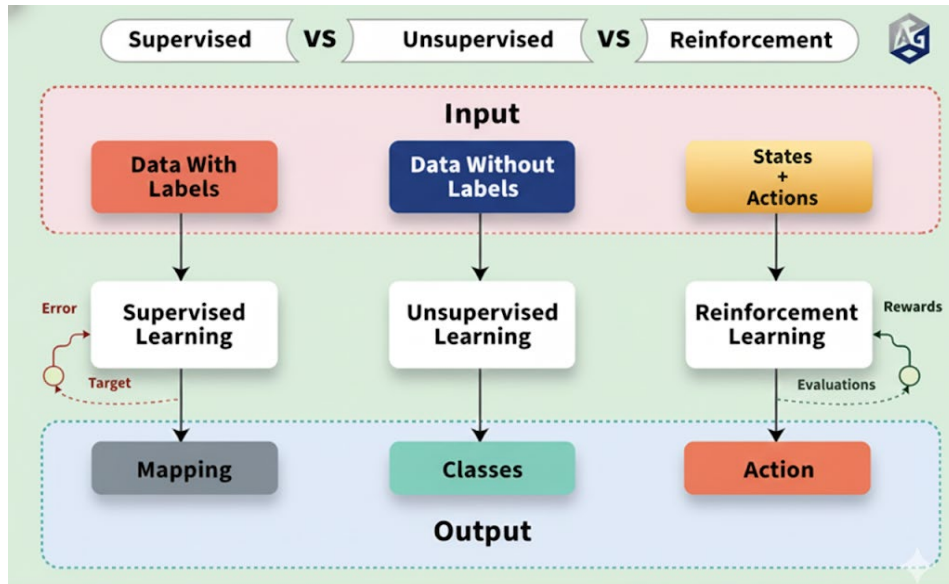
### ٤,١ التعلّم الآلي (Machine Learning)

يُعدّ التعلّم الآلي (Machine Learning – ML) أحد الفروع الأساسية للذكاء الصناعي، كما يتقاطع مع مجال علوم الحاسب، ويهدف بشكل عام إلى تمكين الأنظمة الحاسوبية من التعلّم والتكيف مع المعطيات دون الحاجة إلى برمجتها بشكل صريح لكل حالة على حدة.

تقوم فكرة التعلّم الآلي على بناء نماذج رياضية وخوارزميات قادرة على تحليل البيانات المدخلة واستخلاص الأنماط والعلاقات الكامنة فيها، ومن ثم تحسين أدائها في تنفيذ المهام المطلوبة اعتماداً على الخبرة المكتسبة من البيانات السابقة.

يعتمد التعلّم الآلي على مبدأ التعلّم من التجربة، حيث يتم تدريب النموذج باستخدام مجموعة من المعطيات بهدف تعظيم دقته وتقليل نسبة الخطأ مع مرور الزمن.

ومع ازدياد حجم البيانات وتعقيدها، أصبحت تقنيات التعلّم الآلي أداة أساسية لمعالجة المشكلات التي يصعب حلّها باستخدام الأساليب البرمجية التقليدية، إذ يكون دور التدخل البشري محدوداً ويقتصر غالباً على إعداد البيانات واختيار النموذج المناسب.



Supervised vs Reinforcement vs Unsupervised



## ١,١,٤ أنواع التعلّم الآلي

يُصنّف التعلّم الآلي، وفقاً لطبيعة البيانات المتاحة وآلية التعلّم، إلى عدة أنواع رئيسية، من أهمها ما يلي:

### ١. التعلّم الخاضع للإشراف (Supervised Learning)

يقصد بالتعلّم الخاضع للإشراف تدريب الخوارزمية على بيانات معنونة (Labeled Data)، أي بيانات تحتوي على المدخلات والمخرجات الصحيحة معاً.

خلال عملية التدريب، تتعلّم الخوارزمية بناء علاقة رياضية أو إحصائية بين المدخلات والمخرجات، بما يمكّنها لاحقاً من التنبؤ بمخرجات جديدة غير معروفة اعتماداً على هذه العلاقة.

في أدبيات التعلّم الآلي، يُستخدم هذا النوع من التعلّم في مشكلتين رئيسيتين هما:

التصنيف (Classification)، الذي يهدف إلى التنبؤ بفئات أو أصناف منفصلة، والانحدار (Regression)، الذي يهدف إلى التنبؤ بقيمة مستمرة. من أشهر الخوارزميات المستخدمة في هذا النوع: الانحدار الخطي (Linear Regression) وآلات المتجهات الداعمة (Support Vector Machines).

### ٢. التعلّم غير الخاضع للإشراف (Unsupervised Learning)

يعتمد التعلّم غير الخاضع للإشراف على بيانات غير معنونة (Unlabeled Data)، أي بيانات تحتوي على المدخلات فقط دون وجود مخرجات صحيحة معروفة مسبقاً.

تهدف الخوارزمية في هذا النوع إلى اكتشاف البنية الداخلية للبيانات والتعرّف على الأنماط أو العلاقات بينها دون توجيه مباشر.

تُستخدم تقنيات هذا النوع من التعلّم عادةً في تحليل البيانات الاستكشافي، واكتشاف التجمعات أو البنى المخفية داخل البيانات.

من أشهر خوارزميات التعلّم غير الخاضع للإشراف خوارزميات العنقدة (Clustering).

### ٣. التعلّم شبه الخاضع للإشراف (Semi-Supervised Learning)

يمثل التعلّم شبه الخاضع للإشراف مزيجاً بين التعلّم الخاضع للإشراف وغير الخاضع للإشراف، حيث يعتمد على بيانات معنونة صحيحة ولكنها غير مكتملة، إلى جانب كمية أكبر من البيانات غير المعنونة. يهدف هذا النوع إلى الاستفادة من البيانات غير المعنونة لتحسين أداء النموذج عند توفر عدد محدود من البيانات المعنونة.

### ٤. التعلّم المعزز (Reinforcement Learning)

يُعرّف التعلّم المعزز (Reinforcement Learning) بأنه نمط من أنماط التعلّم الآلي يعتمد على التفاعل المستمر بين النموذج وبيئة معينة. يتعلّم النموذج من خلال التجربة والخطأ، حيث يحصل على مكافآت عند اتخاذ قرارات صحيحة وعقوبات عند اتخاذ قرارات غير مناسبة.

يهدف هذا الأسلوب إلى تعظيم مجموع المكافآت على المدى الطويل، مما يؤدي إلى تحسين أداء النموذج تدريجياً.

### ٤,٢ مفهوم الانحدار (Regression)

يُعدّ الانحدار (Regression) أحد المفاهيم الأساسية في الإحصاء والتعلّم الآلي، ويُستخدم كأداة تحليلية ونمذجة تنبؤية تهدف إلى دراسة العلاقة بين متغير تابع (Dependent Variable) ومتغير واحد أو أكثر من المتغيرات المستقلة (Independent Variables).

يركّز تحليل الانحدار على توصيف هذه العلاقة بشكل رياضي، بما يسمح بفهم تأثير المتغيرات المستقلة على المتغير التابع، إضافةً إلى استخدام النموذج الناتج للتنبؤ بقيم مستقبلية.

في سياق التعلّم الآلي، يُصنّف الانحدار ضمن مسائل التعلّم الخاضع للإشراف، حيث يتم تدريب النموذج على بيانات معنونة تحتوي على القيم الصحيحة للمتغير التابع.

ويتميّز الانحدار عن التصنيف بكونه يتعامل مع مخرجات مستمرة، في حين يركّز التصنيف على مخرجات فئوية أو منفصلة.

## ٤.٢.١ مفهوم تحليل الانحدار

حليل الانحدار هو أسلوب إحصائي يُستخدم لتقدير العلاقة بين المتغيرات، وقياس مدى قوة واتجاه هذا التأثير.

يتطلب بناء نموذج انحدار تحديد العوامل المؤثرة في المتغير التابع، ثم صياغة علاقة رياضية تعبر عن هذه التأثيرات. ويساعد هذا التحليل في تفسير البيانات، واكتشاف الأنماط، واتخاذ قرارات مبنية على أسس كمية.

يعتمد نموذج الانحدار بشكل عام على تقليل الخطأ بين القيم الحقيقية والقيم المتنبأ بها، وذلك باستخدام دوال رياضية يتم تحسين معاملاتها أثناء عملية التدريب.

وتمثل هذه المعاملات وزن تأثير كل متغير مستقل على المتغير التابع.

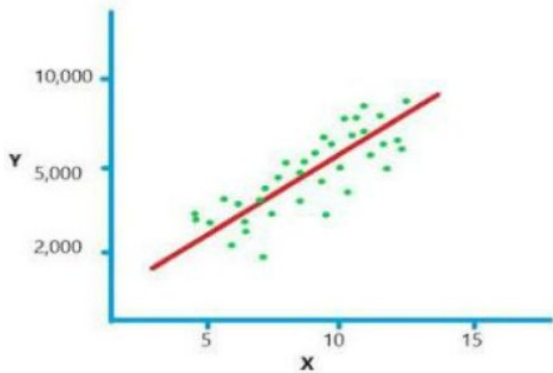
## ٤.٢.٢ أنواع الانحدار

توجد عدة أنواع من تقنيات الانحدار، تختلف باختلاف طبيعة العلاقة بين المتغيرات وبحسب شكل البيانات المستخدمة.

ويُعد اختيار نوع الانحدار المناسب خطوة أساسية لضمان فعالية النموذج ودقته.

### ١. الانحدار الخطي (Linear Regression)

هو أبسط أشكال الانحدار وأكثرها شيوعاً، ويقوم بتقدير العلاقة التي تربط بين متغيرين أحدهما المتغير التابع والآخر المتغير المستقل، وتكون العلاقة التي تربطهما علاقة خطية، وينتج من هذا النموذج معادلة خطية يمكن استخدامها لتفسير العلاقة بينهما أو لتقدير قيمة المتغير التابع عند معرفة قيمة المتغير المستقل.



وتكون العلاقة الرياضية بالشكل التالي:

حيث:  $Y_i$  = المتغير التابع،  $X_i$  = المتغير المستقل،  $\epsilon_i$  = الخطأ العشوائي،  $B_0$  = قيمة ثابتة تعبر عن قيمة المتغير التابع عندما يكون المتغير المستقل مساوياً للصفر،  $B_1$  = هو ميل الخط المستقيم الذي يمثل العلاقة.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

## • الانحدار الخطي المتعدد (Multiple Linear Regression)

يستخدم الانحدار الخطي المتعدد في حال كان لدينا متغيران مستقلان أو أكثر، حيث يوجد معادلة خطية تربط بين المتغير التابع وعدة متغيرات مستقلة شريطة أن تكون المتغيرات غير مرتبطة خطياً، وهو الأكثر فاعلية وواقعية من استخدام متغير مستقل واحد.

وتكون العلاقة الرياضية بالشكل التالي:

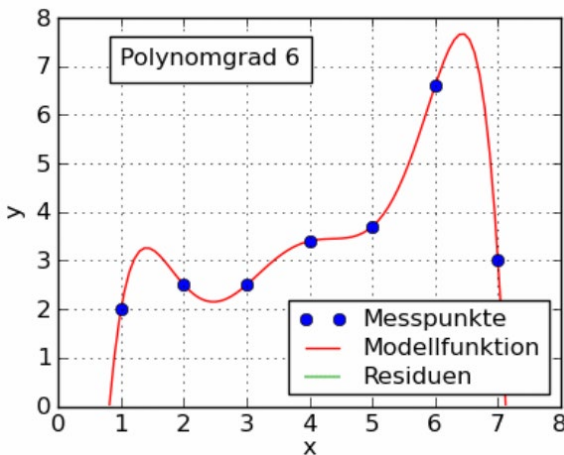
$$y_i = \beta_0 + \sum_{i=1}^n \beta_i x_i + \varepsilon_i$$

حيث:  $Y_i$  = المتغير التابع،  $X_i$  = المتغيرات المستقلة،  $E_i$  = الخطأ العشوائي،  $B_0$  = قيمة ثابتة تعبر عن قيمة المتغير التابع عندما يكون المتغير المستقل مساوياً للصفر،  $B_i$  = هو المعامل المرتبط بالمتغير،  $n$  = تمثل عدد المتغيرات المستقلة (نجد أنه عندما  $n=1$  نعود للانحدار الخطي البسيط).

## ٢. الانحدار اللاخطي (Nonlinear Regression)

### • الانحدار متعدد الحدود (Polynomial Regression)

هو أحد أشكال تحليل الانحدار ويتم فيه نمذجة العلاقة بين المتغيرات المستقلة والمتغيرات التابعة في كثير حدود من الدرجة



$n$ ، والانحدار الخطي هو في الأساس متعدد الحدود من الدرجة الأولى لكن ازدياد درجة الحدود تعطينا الشكل المنحني الأكثر ملاءمة للبيانات (كما نعلم الدرجة الأولى تعبر عن الخط المستقيم)، ولا يتطّب أن تكون العلاقة بين المتغيرات المستقلة والتابعة خطية، وهذا أيضاً أحد الاختلافات بين الانحدار الخطي و متعدد الحدود.

وتكون العلاقة الرياضية بالشكل التالي:

$$y_i = \beta_0 + \sum_{i=1}^n \beta_i x_i^n + \varepsilon_i$$

## ٤.٣ خوارزميات التعلم الآلي المستخدمة في الانحدار

### ٤.٣.١ الانحدار الخطي (Linear Regression)

يُعدّ الانحدار الخطي (Linear Regression) من أبسط وأكثر خوارزميات تعلم الآلة شيوعاً، وينتمي إلى فئة التعلم بإشراف (Supervised Learning)، ويُستخدم في مشاكل الانحدار (Regression) التي يكون فيها المتغير الهدف ذا طبيعة مستمرة.

تقوم هذه الخوارزمية على افتراض وجود علاقة خطية تربط بين المتغير التابع والمتغير أو المتغيرات المستقلة، حيث يتم تمثيل هذه العلاقة من خلال معادلة رياضية خطية.

تعتمد خوارزمية الانحدار الخطي على إيجاد أفضل خط مستقيم يعبر عن العلاقة بين المتغيرات، وذلك من خلال تقليل مقدار الخطأ بين القيم الحقيقية والقيم المتنبأ بها، وغالباً ما يتم ذلك باستخدام طريقة المربعات الصغرى (Least Squares). يُستخدم هذا النموذج بشكل واسع لسهولة تفسيره وبساطة بنيته الرياضية، حيث يتيح فهم تأثير كل متغير مستقل على المتغير التابع بشكل مباشر.

### ٤.٣.٢ الغابات العشوائية (Random Forest Regressor)

تُعدّ خوارزمية الغابات العشوائية (Random Forest) إحدى خوارزميات التعلم التجميعي (Ensemble Learning) التي تعتمد على بناء مجموعة من أشجار القرار واستخدامها بشكل جماعي لإجراء التنبؤ.

تنتمي هذه الخوارزمية إلى التعلم بإشراف، وتُستخدم في كلّ من مشاكل التصنيف والانحدار.

في حالة الانحدار، تقوم الغابات العشوائية ببناء عدد كبير من أشجار القرار، حيث يتم تدريب كل شجرة على عينة عشوائية من البيانات مع اختيار عشوائي لجزء من المتغيرات في كل انقسام.

يتم الحصول على الناتج النهائي من خلال حساب متوسط مخرجات جميع الأشجار، مما يساعد على تقليل التباين وزيادة استقرار النموذج مقارنة باستخدام شجرة قرار واحدة.

تمتاز هذه الخوارزمية بقدرتها على التعامل مع العلاقات غير الخطية والتفاعلات المعقدة بين المتغيرات دون الحاجة إلى افتراضات مسبقة حول شكل العلاقة.

### ٤.٣.٣ خوارزمية XGBoost (Extreme Gradient Boosting)

تُعدّ خوارزمية XGBoost واحدة من أكثر خوارزميات التعزيز المتدرّج (Gradient Boosting) تطوراً وكفاءةً ، وتنتمي إلى فئة التعلّم بإشراف، وتُستخدم بشكل واسع في مسائل الانحدار والتصنيف.

تعتمد هذه الخوارزمية على مبدأ بناء النماذج بشكل تسلسلي، حيث يتم في كل مرحلة تدريب نموذج جديد يهدف إلى تصحيح الأخطاء التي نتجت عن النماذج السابقة.

تعتمد XGBoost على أشجار القرار كنماذج أساسية، إلا أنها تتميز بآليات تحسين متقدمة تساهم في زيادة سرعة التدريب وتقليل مشكلة الإفراط في التعلّم (Overfitting) ، مثل التنظيم (Regularization) والتحكم في تعقيد النموذج.

كما تتميز بقدرتها العالية على التعامل مع البيانات كبيرة الحجم والمعقدة، مما يجعلها من الخوارزميات المفضّلة في تطبيقات التنبؤ ذات الأداء العالي.

### ٤.٣.٤ انحدار متجهات الدعم (Support Vector Regression – SVR)

هو أحد خوارزميات تعلّم الآلة من نوع التعلّم بإشراف “Supervised” ، ويُستخدم في مشاكل الانحدار “Regression” للتنبؤ بالقيم المستمرة.

تعتمد خوارزمية SVR على مبدأ إيجاد دالة انحدار تُحقّق أقلّ انحراف ممكن عن القيم الحقيقية ضمن هامش خطأ محدّد يُعرف بـ  $\epsilon$  (إبسيلون)، مع السعي في الوقت نفسه إلى تعظيم الهامش وتقليل تعقيد النموذج.

تقوم خوارزمية SVR ببناء نموذج يعتمد على مجموعة محدودة من نقاط البيانات تُسمّى متجهات الدعم “Support Vectors”، وهي النقاط الأكثر تأثيراً في تحديد شكل دالة الانحدار.

بخلاف أساليب الانحدار التقليدية التي تهدف إلى تقليل مجموع الأخطاء، يركّز SVR على تجاهل الأخطاء الصغيرة الواقعة ضمن هامش  $\epsilon$  وعدم احتسابها ضمن دالة الخسارة، ممّا يزيد من قدرة النموذج على مقاومة الضجيج وتحقيق تعميم أفضل.

يمكن لخوارزمية SVR التعامل مع العلاقات الخطيّة وغير الخطيّة بين المتغيّرات، وذلك من خلال استخدام دوال النواة “Kernel Functions” التي تسمح بتحويل البيانات إلى فضاء أعلى أبعاداً. من أشهر دوال النواة المستخدمة: النواة الخطيّة، النواة متعدّدة الحدود، ونواة دالة الأساس الشعاعي (RBF)

يتميّز SVR بقدرته على العمل بكفاءة مع البيانات ذات الأبعاد العالية وتقديم نتائج دقيقة في مسائل الانحدار المعقّدة.

## الفصل الخامس

## ٥-المقاربة المقترحة ومنهجية العمل

### ٥,١ مقدمة

تعتمد هذه الدراسة على مقاربة منهجية قائمة على استخدام تقنيات التعلم الآلي لمعالجة مشكلة التنبؤ باستهلاك الطاقة في المباني، وذلك من خلال استثمار البيانات الزمنية المتاحة وبناء نماذج انحدار قادرة على تمثيل السلوك الفعلي للاستهلاك.

وتركّز المقاربة المقترحة على تحويل البيانات الخام إلى تمثيل مناسب للتعلم، ثم تدريب مجموعة من النماذج ومقارنتها ضمن إطار موحد يضمن عدالة التقييم وموثوقية النتائج.

تم تصميم منهجية العمل بحيث تسير عبر مراحل متتابعة ومنفصلة بوضوح، تبدأ بالمعالجة المسبقة للبيانات، يليها استخراج السمات ذات الدلالة الزمنية، ثم تقسيم البيانات وتطبيعها، وصولاً إلى بناء نماذج التعلم الآلي وتقييم أدائها. ويهدف هذا التسلسل إلى ضمان الاستفادة المثلى من المعلومات المتاحة في البيانات، مع الحفاظ على الطبيعة الزمنية للمشكلة وعدم الإخلال بتتابعها.

ولا تقتصر المقاربة المتبعة على الاعتماد على نموذج واحد، بل تشمل مقارنة عدة نماذج انحدار تعتمد مقاربات مختلفة في التعلم، مما يسمح بتحليل سلوك كل نموذج وفهم نقاط القوة والضعف الخاصة به.

ويسهم هذا الأسلوب في اختيار النموذج الأكثر ملاءمة للتنبؤ، إضافة إلى توفير رؤية أعمق حول تأثير السمات المختلفة في عملية التنبؤ.

### ٥,٢ المعالجة المسبقة للبيانات :

في هذه المرحلة، تم إجراء مجموعة من الإجراءات الأساسية على البيانات الخام بهدف جعلها صالحة للاستخدام في نماذج التعلم الآلي.

شملت هذه الإجراءات التحقق من سلامة السجلات الزمنية والتأكد من عدم وجود قيم مفقودة أو غير منطقية قد تؤثر على عملية التدريب.



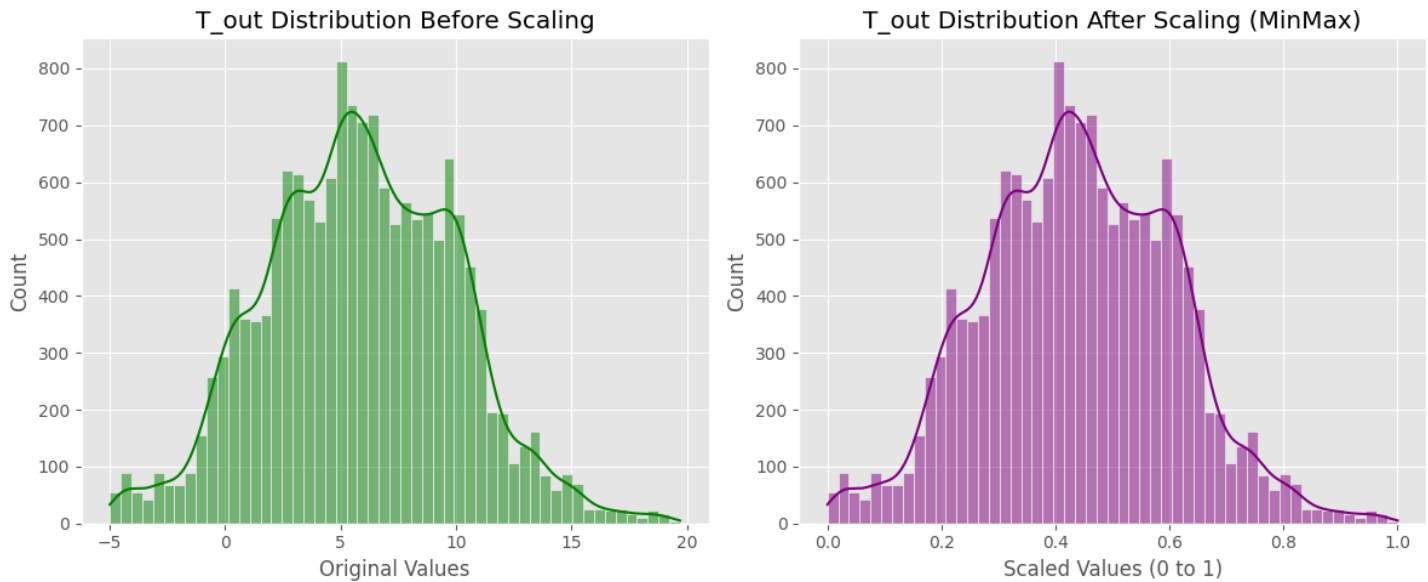
وقد أظهر الفحص المبدئي أن البيانات منظمة زمنياً ولا تحتوي على قيم مفقودة، مما قلّل الحاجة إلى تطبيق عمليات تنظيف معقّدة.

وبناءً على ذلك، اقتصر التنظيف على حذف بعض الأعمدة غير المؤثرة على عملية التنبؤ، بهدف تبسيط البيانات وتحسين كفاءة النماذج، مع الحفاظ الكامل على الترتيب الزمني للسجلات، نظراً للطبيعة الزمنية للمشكلة وعدم جواز الإخلال بتتابعها.

بعد ذلك، تم توحيد مقاييس المتغيرات العددية باستخدام تقنيات التطبيع، وذلك بسبب اختلاف نطاق القيم ووحدات القياس بين المتغيرات المختلفة.

يهدف هذا الإجراء إلى منع المتغيرات ذات القيم الكبيرة من التأثير غير المتوازن على عملية التعلم، إضافةً إلى تحسين استقرار النماذج وتسريع عملية التدريب، خصوصاً عند استخدام نماذج حساسة لمقياس البيانات مثل **Support Vector Regression**.

يوضح الشكل (٢) مثالاً على توزيع أحد المتغيرات العددية قبل وبعد تطبيق خوارزمية **Min-Max Scaling**، حيث تم تحويل القيم إلى مجال موحد بين (٠ و ١) مع الحفاظ على شكل التوزيع الإحصائي للبيانات دون تشويه خصائصها الأصلية.



الشكل (٢)

## ٥,٣ هندسة السمات (Feature Engineering)

نظراً للطبيعة الزمنية لبيانات استهلاك الطاقة، تم التركيز في هذه المرحلة على استخراج سمات مشتقة تساعد النماذج على التقاط الأنماط الزمنية الكامنة في البيانات.

شملت هذه السمات متغيرات الإزاحة الزمنية (Time Lag Features) التي تعبر عن قيم الاستهلاك السابقة، مما يسمح للنماذج بالاستفادة من السلوك التاريخي في عملية التنبؤ.

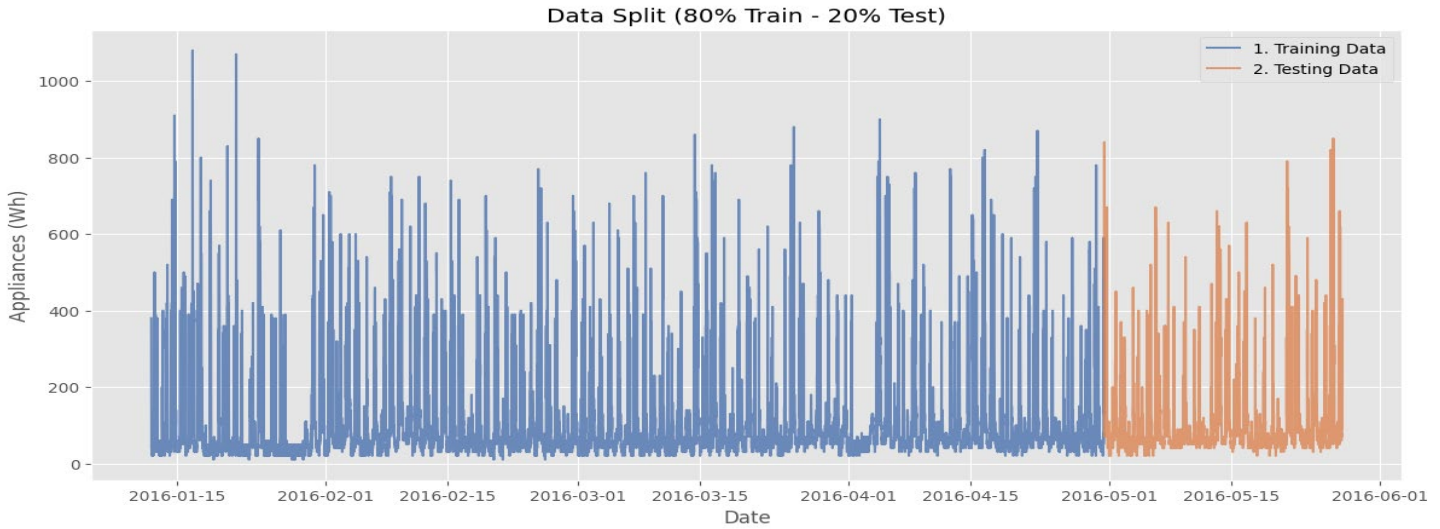
كما تم استخدام المتوسطات المتحركة (Rolling Mean) لتمثيل الاتجاه العام للاستهلاك خلال فترات زمنية قصيرة، الأمر الذي يساهم في تقليل تأثير التقلبات اللحظية والضجيج في البيانات. بالإضافة إلى ذلك، تم تمثيل الخصائص الزمنية الدورية باستخدام الترميز الدائري (Circular Encoding)، بما يضمن الحفاظ على الطبيعة الدورية للوقت دون إدخال انقطاعات مصطنعة في البيانات.

## ٥,٤ تقسيم البيانات وتطبيعها :

بعد الانتهاء من تجهيز السمات، تم تقسيم البيانات إلى مجموعتي تدريب واختبار وفق تسلسل زمني منطقي. حيث خُصص الجزء الأول من السلسلة الزمنية لتدريب النماذج، بينما استُخدم الجزء اللاحق لاختبار قدرتها على التنبؤ بقيم غير مرئية سابقاً.

يضمن هذا الأسلوب محاكاة سيناريو التنبؤ الواقعي، ويمنع تسرب المعلومات من بيانات المستقبل إلى مرحلة التدريب.

تم تطبيق عملية التطبيع باستخدام خوارزمية Min-Max Scaling، حيث تم حساب معاملات التطبيع اعتماداً على بيانات التدريب فقط، ثم تطبيق نفس التحويلات على بيانات الاختبار، وذلك بهدف توحيد نطاق القيم العددية للسمات ومنع تسريب أي معلومات من بيانات الاختبار إلى مرحلة التعلم.



الشكل (٣)

يوضح الشكل (٣) آلية تقسيم البيانات الزمنية إلى مجموعة تدريب (٨٠٪) ومجموعة اختبار (٢٠٪) مع الحفاظ على التسلسل الزمني لسلسلة استهلاك الطاقة.

## ٥,٥ بناء نماذج التعلم الآلي :

في هذه المرحلة، تم بناء مجموعة من نماذج التعلم الآلي الخاضعة للإشراف لمعالجة مشكلة التنبؤ باعتبارها مسألة انحدار.

تم تدريب كل نموذج بشكل مستقل باستخدام نفس مجموعة السمات وبيانات التدريب، مما يسمح بإجراء مقارنة عادلة بين النماذج المختلفة.

تعتمد هذه النماذج على آليات تعلم متنوعة، حيث تمثل بعض الخوارزميات العلاقات الخطية بين المتغيرات ، في حين تعتمد خوارزميات أخرى على نمذجة العلاقات غير الخطية، الأمر الذي يتيح دراسة تأثير اختلاف بنية النموذج على دقة التنبؤ.

## ٥,٥,١ النماذج المستخدمة في التنبؤ :

في إطار المقاربة المقترحة، تم اعتماد مجموعة من نماذج التعلم الآلي الخاضعة للإشراف لمعالجة مسألة التنبؤ باستهلاك الطاقة، باعتبارها مشكلة انحدار تهدف إلى تقدير قيم مستمرة عبر الزمن. تم اختيار هذه النماذج لتمثل فئات مختلفة من خوارزميات الانحدار، وذلك بهدف دراسة أدائها ضمن نفس بيئة البيانات ونفس شروط المعالجة.

تم استخدام نموذج الانحدار الخطي (Linear Regression) كنموذج مرجعي أساسي، لما يتميز به من بساطة واعتماده على فرضية العلاقة الخطية بين المتغيرات، مما يسمح بتقييم مدى كفاية النماذج الخطية في تمثيل سلوك الاستهلاك.

كما تم اعتماد نموذج الغابات العشوائية (Random Forest Regressor) لتمثيل النماذج المعتمدة على أشجار القرار، وذلك لقدرتها على نمذجة العلاقات غير الخطية والتعامل مع التفاعلات المعقدة بين السمات دون الحاجة إلى افتراض شكل رياضي محدد للعلاقة.

بالإضافة إلى ذلك، تم استخدام نموذج آلات المتجهات الداعمة للانحدار (SVR) ، والذي يُعد من النماذج المعتمدة على تعظيم الهامش، ويُستخدم لقياس قدرة هذا النوع من الخوارزميات على تقديم تنبؤات دقيقة عند التعامل مع بيانات ذات أبعاد متعددة.

وأخيراً، تم اعتماد نموذج XGBoost بوصفه نموذج التعزيز التدريجي المتقدم، وذلك لما يُعرف عنه من كفاءة عالية في مسائل الانحدار وقدرته على تحسين الأداء التنبؤي من خلال التعلم التراكمي وتصحيح الأخطاء في المراحل المتتالية.

تم تدريب جميع النماذج باستخدام نفس مجموعة البيانات ونفس إجراءات المعالجة المسبقة، مع الحفاظ على نفس تقسيم بيانات التدريب والاختبار، وذلك لضمان عدالة المقارنة بين النماذج وعزل تأثير بنية النموذج عن باقي العوامل.

## ٥,٦ تقييم النماذج واختيار النموذج الأفضل :

بعد الانتهاء من تدريب النماذج، تم تقييم أدائها باستخدام مجموعة بيانات الاختبار، وذلك من خلال مقارنة القيم المتوقعة بالقيم الحقيقية لاستهلاك الطاقة. أتاح هذا التقييم تحليل سلوك الأخطاء ودراسة قدرة كل نموذج على التعميم والتنبؤ بقيم مستقبلية غير مرئية أثناء التدريب.

وبالاعتماد على نتائج هذا التقييم، تمت مقارنة النماذج من حيث دقة التنبؤ واستقرار الأداء، ليتم اختيار النموذج الأكثر كفاءة واعتماده كنموذج نهائي في مرحلة التنبؤ

## الفصل السادس

## ٦- النتائج و تحليلها

### ٦,١ مقدمة :

يعرض هذا الفصل النتائج التي تم التوصل إليها بعد تدريب واختبار نماذج التعلم الآلي المستخدمة في التنبؤ باستهلاك الطاقة في المباني، وذلك بالاعتماد على مجموعة بيانات **Appliances Energy Prediction Dataset**. كما يتناول الفصل تحليل أداء النماذج المختلفة باستخدام مجموعة من المقاييس الإحصائية المناسبة، مع مناقشة النتائج ومقارنتها فيما بينها لتحديد النموذج الأكثر كفاءة في تمثيل أنماط استهلاك الطاقة.

إضافةً إلى ذلك، يتم تسليط الضوء على أبرز الملاحظات المستخلصة من النتائج، إلى جانب مناقشة التحديات والقيود المحتملة المرتبطة بعملية التنبؤ.

### ٦,٢ مقاييس تقييم الأداء :

تم استخدام المقاييس التالية لتقييم أداء نماذج التنبؤ باستهلاك الطاقة في المباني:

#### • متوسط الخطأ المطلق: (MAE – Mean Absolute Error)

يقيس متوسط القيمة المطلقة للفروق بين القيم الحقيقية والقيم المتنبأ بها، ويعكس مقدار الخطأ المتوسط الذي يرتكبه النموذج في التنبؤ، حيث تشير القيم الأصغر إلى أداء أفضل.

#### • الجذر التربيعي لمتوسط مربع الخطأ: (RMSE – Root Mean Squared Error)

يقيس مدى تباعد القيم المتنبأ بها عن القيم الحقيقية، مع إعطاء وزن أكبر للأخطاء الكبيرة، ويُستخدم لتقييم دقة النموذج وحساسيته للأخطاء الشاذة.

#### • معامل التحديد: (R<sup>2</sup> Score – Model Accuracy)

يمثل نسبة التباين في المتغير الهدف التي يمكن للنموذج تفسيرها اعتماداً على المتغيرات المدخلة، ويُستخدم كمؤشر عام على جودة النموذج، حيث تشير القيم الأقرب إلى الواحد الصحيح إلى أداء أفضل.

## ٦,٣ نتائج النماذج :

في البداية حصلنا على النتائج التالية بعد تجريب عدة نماذج:

### Linear Regression

تم تنفيذ نموذج الانحدار الخطي (Linear Regression) باستخدام لغة البرمجة Python ضمن بيئة Jupyter Notebook، حيث جرى تدريب النموذج واختباره على مجموعة البيانات بعد المعالجة المسبقة.

يوضح الشكل (٤) التالي مثالاً على مخرجات تنفيذ النموذج وقيم مقاييس التقييم المعتمدة.

```
*** --- Linear Regression Baseline Results ---  
MAE (Mean Absolute Error): 26.8723 Wh  
RMSE (Root Mean Squared Error): 54.1835  
R2 Score (Model Accuracy): 0.6412
```

(٤) الشكل

في البداية، تم اعتماد نموذج الانحدار الخطي (Linear Regression) كنموذج مرجعي (Baseline) لتقييم أداء نماذج التنبؤ باستهلاك الطاقة في المباني.

وقد أظهرت النتائج المحصلة أن النموذج حقق قيمة  $MAE = 26.8723 \text{ Wh}$ ، مما يشير إلى أن متوسط الخطأ المطلق بين القيم الحقيقية والمتنبأ بها يبلغ نحو ٢٧ واط-ساعة.

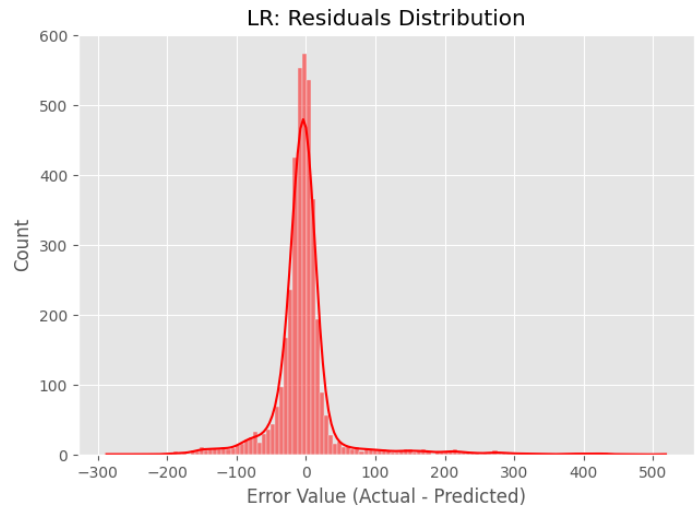
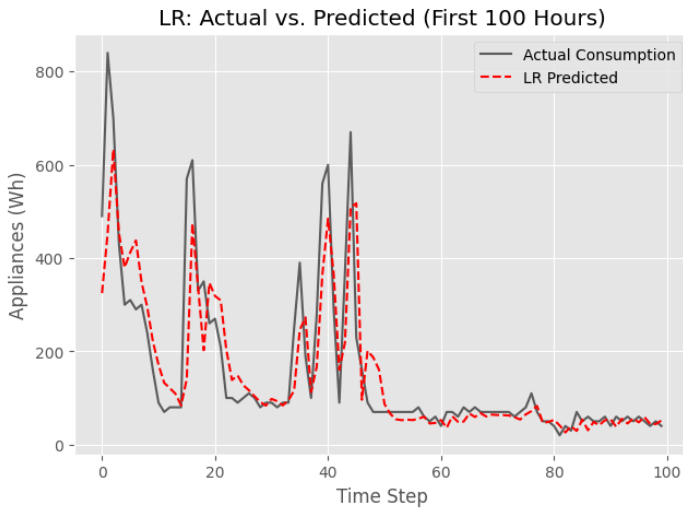
كما بلغت قيمة  $RMSE = 54.1835 \text{ Wh}$ ، وهي أعلى نسبياً من  $MAE$ ، ما يدل على تأثر النموذج بالأخطاء الكبيرة في بعض الفترات الزمنية.

أما معامل التحديد  $R^2 = 0.6412$ ، فيعكس قدرة متوسطة للنموذج على تفسير التباين في بيانات استهلاك الطاقة.

يوضح الشكل (Actual vs. Predicted) مقارنة القيم الحقيقية لاستهلاك الطاقة مع القيم المتنبأ بها خلال أول ١٠٠ ساعة. يمكن ملاحظة أن نموذج الانحدار الخطي ينجح في تتبع الاتجاه العام للاستهلاك، إلا أنه يعاني من صعوبة في تمثيل القمم والانخفاضات الحادة، حيث يظهر فرق واضح بين القيم الحقيقية والمتوقعة في فترات الذروة.

كما يبيّن الشكل (Residuals Distribution) توزيع البواقي (Residuals) ، حيث تتركز الأخطاء حول الصفر مع وجود انتشار ملحوظ على الجانبين، مما يشير إلى أن النموذج لا يعاني من انحياز واضح، لكنه يفتقر إلى القدرة على التقاط العلاقات غير الخطية المعقدة في البيانات.

وتعكس هذه النتائج محدودية نموذج الانحدار الخطي في تمثيل السلوك الديناميكي لاستهلاك الطاقة، مما يبرر الحاجة إلى استخدام نماذج أكثر تقدماً في المراحل اللاحقة.



## Random Forest

تم تطبيق نموذج الغابة العشوائية (Random Forest) كأحد نماذج التعلم الآلي المستخدمة في هذه الدراسة، حيث جرى تدريب النموذج واختباره باستخدام لغة البرمجة Python ضمن بيئة Jupyter Notebook ، وعلى نفس مجموعة البيانات بعد تطبيق مراحل المعالجة المسبقة وهندسة السمات.

يوضح الشكل (٥) التالي مخرجات تنفيذ النموذج، بما في ذلك قيم مقاييس التقييم الإحصائية المعتمدة لقياس دقة التنبؤ.

```
*** --- Random Forest Results ---
MAE (Mean Absolute Error): 28.4456 Wh
RMSE (Root Mean Squared Error): 58.7501
R2 Score (Model Accuracy): 0.5782
```

(٥) الشكل



بالاستناد إلى نتائج التنفيذ السابقة، تم تحليل أداء نموذج الغابة العشوائية (Random Forest) في التنبؤ باستهلاك الطاقة في المباني.

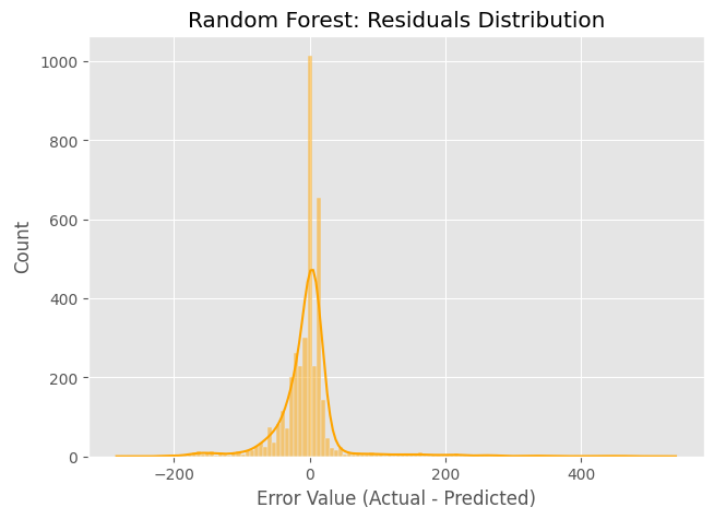
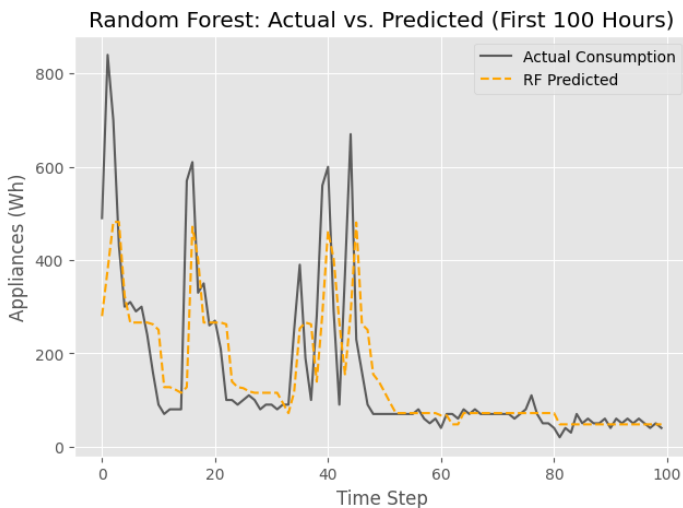
أظهرت النتائج أن النموذج حقق قيمة  $MAE = 28.4456 \text{ Wh}$ ، وقيمة  $RMSE = 58.7501 \text{ Wh}$ ، إضافةً إلى معامل تحديد  $R^2 = 0.5782$ ، وهي قيم تشير إلى أداء أقل مقارنةً بنموذج الانحدار الخطي المستخدم كنموذج مرجعي.

يوضح الشكل (Actual vs. Predicted) مقارنة القيم الحقيقية لاستهلاك الطاقة مع القيم المتنبأ بها خلال أول ١٠٠ ساعة. يمكن ملاحظة أن نموذج Random Forest ينجح في تمثيل الاتجاه العام للاستهلاك، إلا أن دقته تتخفف في فترات الذروة، حيث تظهر فروق واضحة بين القيم الحقيقية والمتوقعة، خصوصاً عند الارتفاعات المفاجئة في الاستهلاك.

كما يبيّن الشكل (Residuals Distribution) توزيع البواقي، حيث تتركز معظم الأخطاء حول الصفر، إلا أن الانتشار الواسع للبواقي على الجانبين يدل على وجود أخطاء كبيرة في بعض الفترات الزمنية.

ويشير ذلك إلى أن النموذج، رغم قدرته على التقاط بعض العلاقات غير الخطية، لم يتمكن من تمثيل السلوك الزمني لاستهلاك الطاقة بدقة عالية ضمن هذا الإعداد.

وبمقارنة نتائج Random Forest مع نموذج الانحدار الخطي، يتضح أن الأداء لم يشهد تحسناً ملحوظاً من حيث قيم  $MAE$  و  $RMSE$  ومعامل التحديد  $R^2$ ، مما يعكس أهمية اختيار النموذج وضبط معاييرها بما يتناسب مع طبيعة البيانات المستخدمة.



# Support Vector Regression – SVR

تم اختبار نموذج آلة المتجهات الداعمة للانحدار (Support Vector Regression – SVR) باستخدام لغة البرمجة Python ضمن بيئة Jupyter Notebook ، حيث جرى تدريبه وتقييمه على نفس مجموعة البيانات بعد تطبيق خطوات المعالجة المسبقة وهندسة السمات.

يوضح الشكل (٦) التالي مخرجات النموذج وقيم مقاييس التقييم الإحصائية المعتمدة.

```
--- SVR Results ---  
** MAE (Mean Absolute Error): 33.7356 Wh  
   RMSE (Root Mean Squared Error): 67.1765  
   R2 Score (Model Accuracy): 0.4485
```

(٦) الشكل

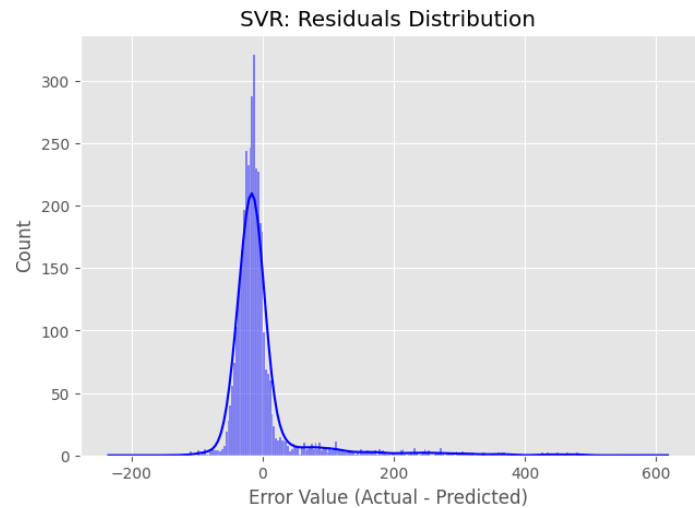
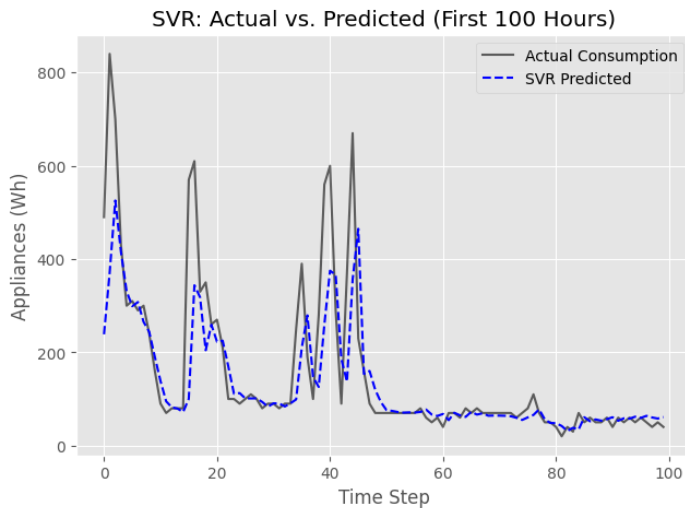
بالاستناد إلى نتائج التنفيذ السابقة، تم تحليل أداء نموذج SVR في التنبؤ باستهلاك الطاقة في المباني.

أظهرت النتائج أن النموذج حقق قيمة  $MAE = 33.7356 \text{ Wh}$  وقيمة  $RMSE = 67.1765 \text{ Wh}$ ، إضافة إلى معامل تحديد منخفض نسبياً بلغ  $R^2 = 0.4485$ ، مما يشير إلى قدرة محدودة للنموذج على تمثيل التباين في بيانات الاستهلاك مقارنة بالنماذج السابقة.

يوضح الشكل (Actual vs. Predicted) مقارنة القيم الحقيقية لاستهلاك الطاقة مع القيم المتنبأ بها خلال أول ١٠٠ ساعة. يمكن ملاحظة أن نموذج SVR يتمكن من تتبع الاتجاه العام للاستهلاك في الفترات المستقرة، إلا أنه يعاني من صعوبة واضحة في تمثيل القمم والتغيرات الحادة، حيث يظهر تفاوت ملحوظ بين القيم الحقيقية والمتوقعة في فترات الذروة.

كما يبين الشكل (Residuals Distribution) توزيع البواقي، حيث تتركز معظم الأخطاء حول الصفر، إلا أن وجود ذيول ممتدة يشير إلى حدوث أخطاء كبيرة في بعض الحالات، مما يعكس محدودية النموذج في التقاط العلاقات غير الخطية المعقدة والسلوك الزمني المتغير لاستهلاك الطاقة.

وبمقارنة نتائج نموذج SVR مع كل من الانحدار الخطي (Linear Regression) و Random Forest، يتضح أن أدائه كان الأضعف من حيث قيم MAE و RMSE و معامل التحديد، الأمر الذي يبرز الحاجة إلى نماذج أكثر مرونة وقدرة على التعامل مع طبيعة البيانات الديناميكية في هذا النوع من التطبيقات.



## XGBoost Regressor

بعد مقارنة أداء النماذج المختلفة، تم اختيار نموذج XGBoost بوصفه النموذج الأفضل في هذه الدراسة، نظراً لتفوقه من حيث دقة التنبؤ واستقراره الإحصائي.

يوضح الشكل التالي النتائج النهائية للنموذج، والتي تعكس الأداء الأعلى مقارنة ببقية النماذج المستخدمة

```
*** Training the FINAL CHAMPION Model (XGBoost)...
--- XGBoost Regressor - FINAL CHAMPION Results ---
MAE (Mean Absolute Error): 25.6668 Wh
RMSE (Root Mean Squared Error): 52.4573
R2 Score (Model Accuracy): 0.6637

Final champion accuracy established: R2 Score: 0.6637
```

(٧) الشكل

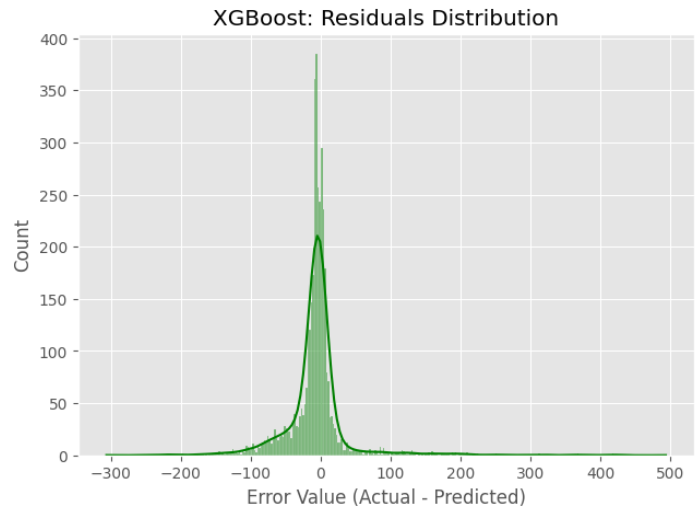
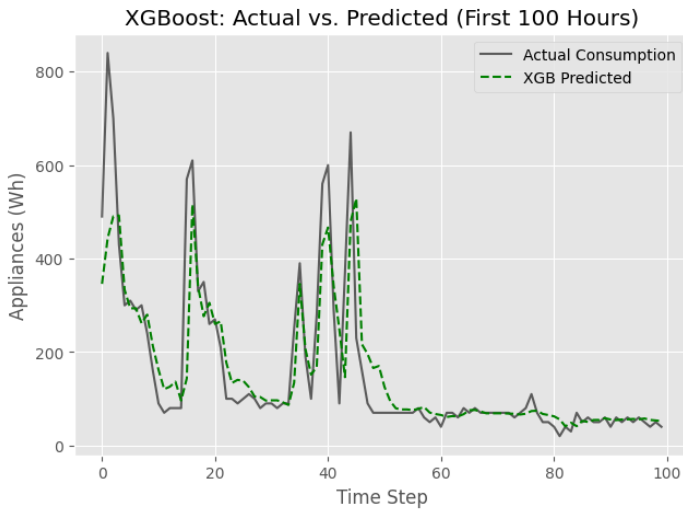
في ضوء نتائج المقارنة، تم اختيار نموذج XGBoost Regressor بوصفه النموذج النهائي والأكثر كفاءة في التنبؤ باستهلاك الطاقة في المباني.

أظهرت النتائج تفوق هذا النموذج على بقية النماذج المستخدمة، حيث حقق أقل قيمة  $MAE = 25.6668 \text{ Wh}$ ، وأدنى قيمة  $RMSE = 52.4573 \text{ Wh}$ ، إضافةً إلى أعلى معامل تحديد بلغ  $R^2 = 0.6637$ ، مما يعكس قدرته العالية على تمثيل التباين في بيانات الاستهلاك بدقة أفضل.

يوضح الشكل (Actual vs. Predicted) مقارنة القيم الحقيقية لاستهلاك الطاقة مع القيم المتنبأ بها خلال أول ١٠٠ ساعة، حيث يُلاحظ أن نموذج XGBoost يتمكن من تتبع السلوك العام للاستهلاك بدقة أعلى مقارنة بالنماذج الأخرى، لا سيما في تمثيل القمم والتغيرات الحادة، مع تقليل واضح للفجوة بين القيم الحقيقية والمتوقعة.

كما يبيّن الشكل (Residuals Distribution) توزيع البواقي، حيث تتركز الأخطاء بشكل أكبر حول الصفر مع انخفاض ملحوظ في انتشارها مقارنة بالنماذج السابقة، مما يشير إلى تحسّن استقرار النموذج وتقليل الأخطاء الكبيرة. ويعكس هذا التوزيع المتوازن قدرة XGBoost على النقاط العلاقات غير الخطية والتفاعلات المعقدة بين المتغيرات بشكل أكثر فاعلية.

وبناءً على المقاييس الإحصائية وتحليل الرسومات، يمكن اعتبار نموذج XGBoost النموذج الأنسب للتنبؤ باستهلاك الطاقة ضمن هذا العمل، مما يبرر اعتماده كنموذج نهائي للمقارنة والتحليل في هذا المشروع



## ٦,٤ تحليل النتائج :

تُظهر النتائج أن نماذج التعلّم الآلي المستخدمة تمتلك قدرة متفاوتة على تمثيل أنماط استهلاك الطاقة في المباني، حيث برزت النماذج التجميعية، وبشكل خاص XGBoost، بقدرة أعلى على التقاط السلوك العام للاستهلاك والتعامل مع التغيرات غير الخطية في البيانات.

ويعكس ذلك ملاءمة هذه النماذج للتعامل مع الطبيعة الديناميكية لاستهلاك الطاقة، والتي تتأثر بعوامل زمنية وتشغيلية متعددة.

كما لوحظ أن النماذج الأبسط، مثل الانحدار الخطي وSVR، تواجه صعوبة في تمثيل القمم والانخفاضات الحادة في الاستهلاك، مما يشير إلى محدودية قدرتها على تمثيل العلاقات المعقدة بين المتغيرات.

في المقابل، ساهمت البنية التجميعية في تقليل تأثير الضجيج وتحسين استقرار التنبؤات، وهو ما انعكس على توزيع البواقي وتقارب القيم المتنبأ بها من القيم الحقيقية.

وبالرغم من التحسن الملحوظ في أداء بعض النماذج، إلا أن وجود أخطاء في فترات الذروة يشير إلى أن استهلاك الطاقة يتأثر بعوامل مفاجئة أو غير ممثلة بشكل كامل في البيانات، مما يفتح المجال أمام تطوير نماذج أكثر تقدماً أو دمج مصادر بيانات إضافية مستقبلاً.

ومع ذلك، تؤكد النتائج أن استخدام تقنيات التعلّم الآلي يمثل أداة فعّالة وموثوقة في دعم أنظمة إدارة الطاقة وتحسين كفاءتها في المباني.

## ٦,٥ مقارنه مع دراسات سابقة :

بالمقارنة مع الدراسة المرجعية

“Data-driven prediction models of energy use of appliances in a low-energy house”

والتي اعتمدت على نماذج إحصائية وتقنيات تعلّم آلي تقليدية مثل الانحدار الخطي المتعدد، وآلة المتجهات الداعمة، والغابة العشوائية، ونماذج تعزيز التدرج، تُظهر نتائج هذا البحث تفوقاً واضحاً في أداء التنبؤ باستخدام نموذج أكثر تطوراً ضمن نفس إطار البيانات.

وعلى الرغم من أن كلا الدراستين تؤكدان فعالية النماذج التجميعية في التنبؤ باستهلاك الطاقة في المباني، فإن النتائج الحالية تُبرز إمكانية تحقيق أداء أعلى عند استخدام خوارزميات محسّنة تعتمد على آليات تعزيز متقدمة وضبط أدق للمعاملات، دون الحاجة إلى تغيير مجموعة البيانات أو إضافة مصادر معلومات جديدة.

ويعكس هذا التفوق قدرة النموذج المستخدم في هذا البحث على تمثيل العلاقات غير الخطية والتفاعلات المعقدة بين المتغيرات بشكل أكثر كفاءة.

ويعزز هذا التقارب في المنهجية، إلى جانب التفوق في النتائج، من القيمة العلمية لهذا العمل باعتباره تطويراً تطبيقياً قائماً على دراسة مرجعية معروفة، ويؤكد أن تحسين بنية النموذج واستراتيجية التعلم يمكن أن يؤدي إلى نتائج أدق وأكثر موثوقية في مجال التنبؤ باستهلاك الطاقة.

## ٦,٦ التحديات والتحسينات المحتملة:

على الرغم من الأداء الجيد الذي حققته نماذج التعلم الآلي في التنبؤ باستهلاك الطاقة، إلا أن هناك بعض التحديات التي ما تزال قائمة، لا سيما في تمثيل فترات الذروة والتغيرات المفاجئة في الاستهلاك، إضافةً إلى التأثير بالعوامل غير الممثلة بشكل كامل ضمن البيانات المتاحة.

كما قد تؤدي طبيعة البيانات الزمنية وسلوك الاستهلاك غير المنتظم في بعض الفترات إلى زيادة الخطأ في التنبؤ.

مستقبلاً، يمكن تحسين النتائج من خلال:

- استخدام نماذج أكثر تطوراً، مثل النماذج الهجينة أو تقنيات التعلم العميق القادرة على تمثيل الاعتماديات الزمنية المعقدة بشكل أدق.
- زيادة حجم البيانات التدريبية أو توسيع فترة الرصد الزمني، مما يساهم في تحسين قدرة النماذج على التعميم وتقليل الأخطاء.
- تحسين أساليب المعالجة المسبقة للبيانات، بما في ذلك اختيار المتغيرات الأكثر تأثيراً وتقليل الضجيج، الأمر الذي ينعكس إيجاباً على استقرار ودقة التنبؤات.

## الفصل السابع

## ٧-الخاتمة و التوصيات

### ٧,١ الخاتمة:

يهدف هذا البحث إلى دراسة وتطوير نماذج ذكية للتنبؤ باستهلاك الطاقة في المباني، اعتماداً على تقنيات التعلم الآلي وتحليل البيانات الزمنية، وذلك باستخدام مجموعة بيانات **Appliances Energy Prediction Dataset**.

تم في هذا العمل تقييم عدة نماذج شملت الانحدار الخطي، آلة المتجهات الداعمة (SVR)، الغابة العشوائية (Random Forest)، ونموذج **XGBoost**، بهدف تحليل قدرتها على تمثيل أنماط الاستهلاك والتغيرات الزمنية في البيانات.

أظهرت النتائج أن النماذج التجميعية، وبشكل خاص نموذج **XGBoost**، حققت أداءً أفضل مقارنة بالنماذج التقليدية، حيث أبدت قدرة أعلى على تمثيل العلاقات غير الخطية وتقليل أخطاء التنبؤ.

كما بينت المقارنة بين النماذج أن النماذج الأبسط تواجه صعوبة في تمثيل القمم والتغيرات الحادة في استهلاك الطاقة، في حين أظهرت النماذج التجميعية استقراراً ودقة أفضل في التنبؤ.

ساهمت مراحل المعالجة المسبقة للبيانات وتنظيمها في تحسين جودة المدخلات المستخدمة في التدريب، مما انعكس إيجاباً على أداء النماذج، رغم وجود بعض التحديات المرتبطة بالطبيعة الديناميكية لاستهلاك الطاقة وتأثره بعوامل متعددة.

وبشكل عام، يبين هذا البحث أن تقنيات التعلم الآلي تمثل أداة فعالة لدعم أنظمة إدارة الطاقة في المباني، وتحسين كفاءتها التشغيلية، والمساهمة في تقليل الهدر الطاقوي وتعزيز الاستدامة.

### ٧,٢ التوصيات :

استناداً إلى النتائج التي تم التوصل إليها في هذا البحث، يمكن تقديم التوصيات التالية:

١. توسيع قاعدة البيانات :يُوصى باستخدام بيانات تمتد لفترات زمنية أطول، أو جمع بيانات من مبانٍ مختلفة، لتعزيز قدرة النماذج على التعميم وتحسين دقة التنبؤ.
٢. استخدام نماذج أكثر تقدماً :يمكن مستقبلاً اعتماد نماذج هجينة أو نماذج تعلم عميق قادرة على تمثيل الاعتماديات الزمنية المعقدة بشكل أدق، خاصة في حالات التغير المفاجئ في الاستهلاك.



٣. تحسين المعالجة المسبقة للبيانات: يُنصح بالتركيز على تقنيات متقدمة لتنقية البيانات، واختيار المتغيرات الأكثر تأثيراً، وتقليل الضجيج، لما لذلك من أثر مباشر على استقرار النموذج ودقته.
٤. اختبار النماذج في بيئات واقعية: من المفيد تطبيق النماذج المقترحة ضمن أنظمة إدارة طاقة حقيقية في المباني، للتحقق من فعاليتها في الظروف التشغيلية الفعلية.
٥. تطوير واجهة استخدام ذكية: يمكن تصميم واجهة تفاعلية تسهل عرض استهلاك الطاقة والتنبؤات المستقبلية، مما يساعد المستخدمين وصناع القرار على الاستفادة العملية من نتائج النموذج.

### ٧,٣ افاق مستقبلية :

- يفتح هذا البحث المجال أمام تطوير أنظمة تنبؤ أكثر تقدماً تعتمد على دمج مصادر بيانات متعددة، مثل بيانات الطقس، الإشغال، وسلوك المستخدمين، مما يساهم في تحسين دقة التنبؤ وواقعيته.
- كما يمكن مستقبلاً توظيف تقنيات الذكاء الاصطناعي القابل للتفسير (Explainable AI) لفهم سلوك النماذج وتحليل أسباب التنبؤات، مما يعزز الثقة في استخدامها ضمن أنظمة إدارة الطاقة الذكية.
- كذلك، يمكن ربط هذه النماذج بأنظمة مراقبة لحظية أو منصات إنترنت الأشياء (IoT) ، لتحقيق تنبؤات آنية تدعم اتخاذ القرار في المباني الذكية وتساهم في تحقيق كفاءة طاقية أعلى واستدامة بيئية أفضل.

# References ٨-المراجع

- [1] Hyndman, Rob J., and George Athanasopoulos. *Forecasting: Principles and Practice*. 3rd edition, OTexts, 2021.
- [2] Brownlee, Jason. *Machine Learning for Time Series Forecasting with Python*. Machine Learning Mastery, 2018.
- [3] Witten, Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal. *Data Mining: Practical Machine Learning Tools and Techniques*. 4th edition, Morgan Kaufmann, 2016.
- [4] Kuhn, Max, and Kjell Johnson. *Applied Predictive Modeling*. Springer, 2013.
- [5] Box, George E.P., Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. *Time Series Analysis: Forecasting and Control*. 5th edition, Wiley, 2015.
- [6] Zhang, Y., Wang, S., and Li, H. "Forecasting energy consumption time series using machine learning techniques." *Energy*, vol. 165, pp. 709-726, 2018.
- [7] Calvillo, Carlos, and Javier García. "Traditional Machine Learning Models for Building Energy Performance Prediction." *Machine Learning Research Journal*, vol. 8, no. 1, pp. 11-25, 2023.
- [8] Sadeghzadeh, Mohammad, et al. "Machine learning for estimation of building energy consumption and performance: a review." *Visual Computing for Industry, Biomedicine, and Art*, 1:15, 2018.
- [9] Chou, Jason, et al. "Smart Buildings Energy Consumption Forecasting using Adaptive Evolutionary Ensemble Learning Models." *arXiv preprint arXiv:2506.11864*, 2025.
- [10] Chen, X., et al. "XGBoost for Regression Predictive Modeling and Time Series." 2022.