

تحليل حركة المرور ضمن المدن الذكية

م. خالد قندججي د. ماهر عباس* د. أسماء شعار*

المخلص

باتت خوارزميات تعلم الآلة جزءاً مهماً ضمن بيئة انترنت الأشياء، حيث تسهم هذه الخوارزميات في تقديم كثير من الخدمات والتطبيقات، وذلك بالاستفادة من معالجة البيانات المتولدة عن هذه الشبكات وتحليلها، تعد المدن الذكية من أهم تطبيقات انترنت الأشياء، التي تحوي بدورها على مجموعة وافرة من التطبيقات الفرعية كتلك التي تخص نظم التنقل الذكية، نظم إدارة مواقف السيارات ودراسة الطقس والتلوث ضمن المدينة وغيرها. تستخدم خوارزميات الانحدار للتنبؤ بالقيم المستمرة حيث تستخدم هذه الخوارزميات في كثير من التطبيقات كالتنبؤ بأسعار الصرف، أسعار المنازل وغيرها. قمنا في هذا البحث بدراسة وتحليل أداء خوارزميات الانحدار المختلفة في مسألة التنبؤ بحركة المرور المستقبلية ضمن المدن الذكية الأمر الذي يساعد في تقديم العديد من الخدمات كإدارة مواقف السيارات وتنظيم الإشارات المرورية وتخفيف الازدحام وغيرها. قمنا في هذا البحث بدراسة وتحليل حركة المرور باستخدام مجموعة بيانات مقدمة من مشروع citypulse والتي تحتوي على عدد من الخواص كسرعة الحركة، وبيانات تلوث الجو، كما وتم دراسة تأثير عامل الزمن من خلال استخدام القراءات المسجلة سابقاً في عملية التحليل وبينت الدراسة أن النماذج المعتمدة على القراءات السابقة حققت دقة أعلى، فيما تفوقت خوارزميتي الانحدار التدريجي والغابات العشوائية على باقي الخوارزميات حيث حققت دقة جيدة من أجل مجموعة البيانات المدروسة.

الكلمات المفتاحية: حركة المرور، التنبؤ، المدن الذكية، تحليل البيانات.

م. خالد قندججي طالب دراسات عليا دكتوراه في كلية الهندسة المعلوماتية جامعة البعث

د. ماهر عباس أستاذ مساعد في كلية الهندسة المعلوماتية جامعة البعث قسم هندسة الشبكات والنظم الحاسوبية.

Smart City Traffic Flow Analysis

Abstract

The use of machine learning algorithms has become increasingly prevalent in the IoT environment, as they offer valuable services and applications by analyzing and interpreting data generated within these networks. Smart cities are among the most important applications of the Internet of Things, which, in turn, contain a wide range of applications. Such as intelligent transportation systems, parking management systems, weather and pollution analyzing systems. Regression algorithms are used to predict continuous values, and these algorithms are applied in many areas, such as predicting exchange rates and house prices. In this research, we analyzed the performance of different regression algorithms to predict smart city traffic in order to enhance various services such as parking area management, traffic signal coordination, and avoiding traffic congestion. We analyze the traffic using attributes like traffic speed and pollution data provided by citypulse project, we also study the effect of use previously recorded values in prediction process, models based on previously recorded values achieved higher accuracy. The study showed that both the gradient boosting and random forest algorithms outperformed the other algorithms and achieved good accuracy for the dataset studied.

KEYWORDS: traffic flow, predicting, smart cities, data analysis.

١ - مقدمة:

تستخدم تقنيات شبكات انترنت الأشياء في شتى مجالات الحياة حيث باتت جزءاً لا يتجزأ في عديد من التطبيقات التي تسهل حياة الأفراد، نذكر من هذه التطبيقات البيوت الذكية (Smart Homes)، المدن الذكية (Smart Cities)، الأنظمة الصحية الذكية (Smart Healthcare) وأنظمة النقل الذكية (ITS (Intelligent Transportation Systems وغيرها العديد من التطبيقات.

تعاني معظم الدول وبشكل متزايد من مشكلة الازدحام المروري وبناءً على تقرير منظمة الصحة العالمية (WHO (World Health Organization) ١,٣ مليون شخص يموتون كل سنة جراء حوادث مرورية وحوالي ٢٠-٥٠ مليون آخرين يعانون من إصابات أو إعاقات جراء هذه الحوادث [1]. يتسبب الازدحام في ساعات الذروة بالإضافة لأعمال البناء وسوء تنظيم عمل إشارات المرور وتوقيتها وغيرها بانتشار التلوث بأنواعه (تلوث الهواء والضجيج) والحوادث المرورية [2].

لحل المشاكل المتعلقة بالازدحام المروري وأسبابها ظهرت نظم التنقل الذكية المعتمدة على شبكات انترنت الأشياء (IoT (Internet of Things من خلال جمع بيانات تخص حالة الطريق والحوادث المرورية لتأمين متطلبات التنقل الآمن من خلال تحديد المسار الأفضل، إخطار السائقين بالحوادث والازدحام المروري، تزويد السائقين بمعلومات تخص مواقف السيارات وحالة الطقس وغيرها... تعتمد هذه النظم على تقنيات الذكاء الصناعي وتحليل البيانات بشكل رئيسي لتحقيق المتطلبات السابقة لتضمينها ضمن نظام الملاحة الخاص بالسيارات والأنظمة المعتمدة على تبادل الرسائل مع مصادر البيانات الأخرى [3].

تعد عملية التنبؤ بحركة المرور ضمن المدن من أهم الطرق المستخدمة ضمن نظم التنقل الذكية وبشكل عام تكون عملية التنبؤ أحادية الخواص univariate عند الاعتماد على

عدد المركبات ضمن موقع معين للتنبؤ أو متعددة الخواص وذلك عند استخدام خواص متعددة كسرعة الحركة، حالة الجو والتلوث وغيرها. [5].

فيما يأتي سنتحدث عن مشكلة وهدف البحث ثم نتطرق لأهمية البحث ومن ثم نستعرض أهم الدراسات في هذا المجال. وفي الفقرات التالية سنتحدث عن نظم التنقل الذكية والسلاسل الزمنية وأهم الخوارزميات المستخدمة للتنبؤ ومجموعة البيانات المستخدمة ونختم بالنتائج والمقترحات والتوصيات المستقبلية.

٢- مشكلة البحث:

تحتوي المدن الذكية على كثير من البيانات التي يتم تسجيلها عن طريق الحساسات مما يجعل تحليل هذه البيانات أمراً هاماً للاستفادة منها [3]، كما وتوجد العديد من العوامل التي تؤثر على حركة المرور ضمن المدينة كسرعة الحركة، التلوث وغيرها. تمتاز بيانات حركة المرور بالدورية [6]، حيث يكون لدينا فترات ذروة وفترات راحة ضمن اليوم الواحد مثلاً مما يجعل من تحليل هذه البيانات وتأثير الخواص المسجلة من قبل المدن الذكية ذو فائدة لتحسين تطبيقات التنقل الذكية.

٣- هدف البحث:

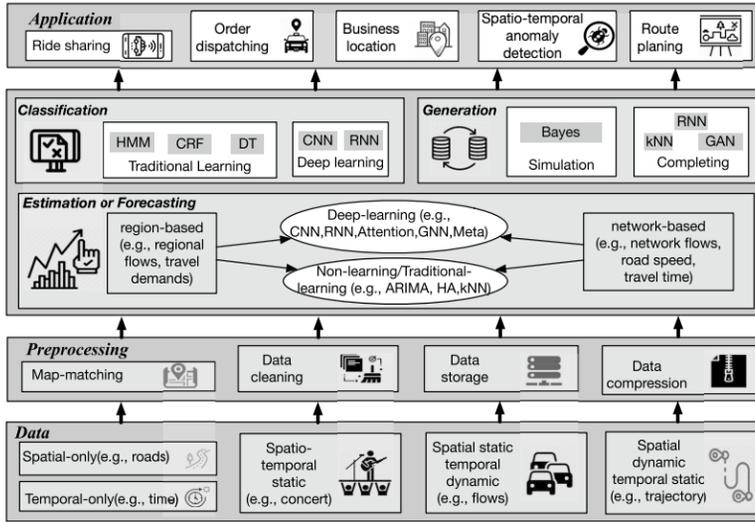
يهدف البحث للتنبؤ بحركة المرور ضمن المدن الذكية ودراسة تأثير بعض الخواص عليها كسرعة الحركة وقيم الانبعاثات الغازية ك (CO_2 , SO_2) التي تمثل حالة التلوث ضمن الجو بشكل مستقل عن القراءات السابقة أي بالاعتماد على القيم اللحظية التي تم قراءتها من الحساسات، بالإضافة لما سبق يهدف البحث أيضاً لدراسة تأثير قراءات الحساسات السابقة بتعبير آخر دراسة تأثير عامل الزمن على عملية التنبؤ وعلى أداء النماذج المستخدمة.

٤- أهمية البحث:

تقدم عملية التنبؤ بحركة المرور العديد من الفوائد التي يمكننا الاستفادة منها في تطبيقات نظم التنقل الذكية منها الضبط الأمثل لزمان إشارات المرور، تنظيم حركة المركبات ضمن المدن وتخفيف الازدحام ومساعدة الجميع للوصول للهدف بأسرع وقت، بالإضافة لضبط مواقف المركبات وتوفير بيئة مناسبة لحركة الجميع بسلام وأمان.

٥- نظم التنقل الذكية (ITS):

يوضح الشكل (1) البنية العامة لنظم التنقل الذكية والذي يتكون من المكونات التالية:



الشكل (١) البنية العامة لنظم التنقل الذكية ITS [4]

1-5 البيانات (Data) [4]:

تتنوع البيانات المستخدمة ضمن نظم التنقل الذكية، فقد تكون هذه البيانات ثابتة كإحداثيات الطرق والمواقع وبعضها ديناميكياً كحركة المرور يتم تجميعها باستخدام شبكات إنترنت

الأشياء أو الحساسات مثلاً. وبشكل عام تتكون البيانات المستخدمة ضمن نظم التنقل الذكية من:

- ١- بيانات مكانية (Spatial Data): يتضمن هذا النوع من البيانات إحداثيات النقاط المهمة ضمن المدينة وإحداثيات الطرق وهي مستقلة عن الزمن.
- ٢- بيانات زمنية (Temporal Data): تتضمن التاريخ والوقت أو الوسمة الزمنية وأيام العطل وهي مستقلة عن القيم المكانية.
- ٣- بيانات ثابتة على مستوى الزمان والمكان (Spatial-temporal Static Data): ومن الأمثلة على هذه البيانات كالمناسبات الاجتماعية والفعاليات ضمن المدينة.
- ٤- بيانات ثابتة على مستوى المكان ومتغيرة الزمن (Spatial Static Temporal Dynamic Data): حيث يتم تمثيل البيانات على شكل سلسلة زمنية (time series) كحركة المرور وسرعة الحركة ضمن طريق ما وحالة الطقس.

2-5 معالجة البيانات (Data Preprocessing) [4]:

تتعرض البيانات لمشاكل عديدة كالنقص، وصول قيم شاذة نتيجة أي خطأ قد يحدث أثناء عملية نقل البيانات ضمن شبكات انترنت الأشياء، لذا لابد من تنظيف البيانات قبل البدء بأي عملية تحليل لها، كما ولابد من توفير بيئة تخزين مناسبة لهذا النوع من البيانات حسب طبيعتها كالبني الموزعة ومنها (Apache spark and Hadoop) التي توفر إمكانيات عديدة كمعالجة تدفقات البيانات، التحليل السريع وسهولة الاستخدام [28].

3-5 طبقة تحليل البيانات [4]:

بشكل عام يوجد نوعين من العمليات التي يمكننا استخدامها على هذه البيانات:

١- التصنيف (Classification): وفيها تستخدم هذه الخوارزميات للتنبؤ بحالة الطريق (مزدحم أم لا).

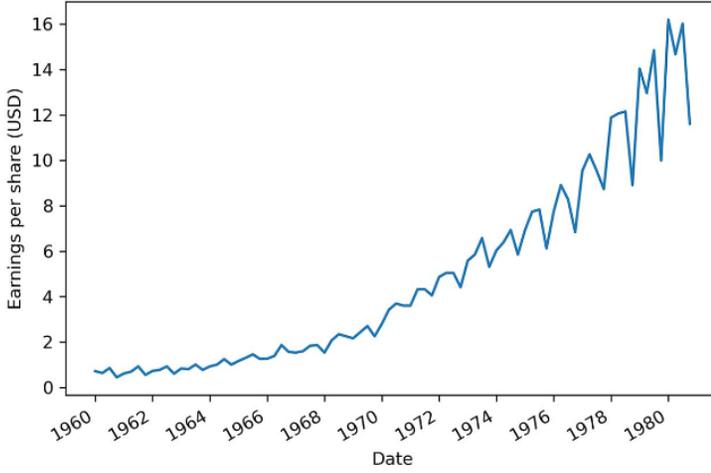
٢- التنبؤ (Prediction): وفيها تستخدم هذا النوع من الخوارزميات للتنبؤ مثلاً بحركة المرور ضمن طريق ما، ونستخدم في هذه الحالة البيانات التي يتم تجميعها على النقاط المثبتة ضمن الطرق. ومثال آخر التنبؤ بزمن تنقل السكان لأجل تحسين عملية النقل باستخدام سيارات الأجرة بحيث يتم نقل السكان بوقت مقبول ودون تأخير.

4-5 طبقة التطبيقات (Application Layer) [4]:

تُساهم عملية تحليل البيانات بتقديم العديد من التطبيقات المفيدة لتساعد في تحسين حركة المرور ضمن المدينة كإكتشاف الشذوذ في حركة المرور [27]، تخطيط الطرق، تموضع سيارة الأجرة ضمن المدينة، تنظيم إشارات المرور وغيرها من الخدمات الأخرى.

٦- السلاسل الزمنية (Time Series):

تُعرف السلسلة الزمنية بأنها مجموعة من نقاط البيانات المرتبة زمنياً، بتعبير آخر هي عبارة عن تسلسل زمني من العينات التي تخص المتحول المدروس [26]، يتم تسجيل هذه النقاط خلال فواصل زمنية متساوية (كل دقيقة، ساعة، يوم، ...)، ولا تعتبر مجموعة النقاط المدروسة سلسلة زمنية إلا إذا تم تسجيل النقاط خلال فواصل زمنية متساوية. تستخدم السلاسل الزمنية في الكثير من التطبيقات اليومية في حياتنا كتسجيل استهلاك الطاقة الكهربائية، مراقبة درجة حرارة الجو، في مجال الاقتصادي وغيرها الكثير.... [26]. يوضح الشكل [2] سلسلة زمنية الإيرادات المالية لمجمع تجاري تم تسجيلها على مدار أرباع السنوات من 1960 وحتى 1980.



شكل (2) السلسلة الزمنية الممثلة لإيرادات مجمع تجاري [26].

لتحليل السلاسل الزمنية بشكل ناجح يجب أن يتم دراسة مركبات هذه السلاسل بشكل مفصّل ولأجل ذلك نستخدم عملية تقسيم السلسلة الزمنية (Decomposition)، والتي تعرف على أنها تحليل إحصائي للسلاسل الزمنية يهدف لفصل مركبات السلسلة الزمنية عن بعضها البعض، تفيد هذه العملية في استخلاص اتجاه (trend) السلسلة والفصلية (seasonality) على سبيل المثال والتي تساعد في فهم طبيعة السلسلة المدروسة. تتكون السلاسل الزمنية بشكل عام من المركبات التالية:

١- قيم السلسلة الزمنية: تمثل السلسلة الزمنية المسجلة.

٢- الاتجاه (Trend): يمثل التغير الحاصل في السلسلة الزمنية ويمكن أن يكون يتصاعد أو يتنازل حسب طبيعة السلسلة المدروسة كما وأننا نرسم خطأ مستقيماً يوضح اتجاه السلسلة الزمنية.

٣- الفصلية أو الموسمية (Seasonality): تلتقط هذه المركبة التغيرات الدورية للسلسلة خلال فواصل زمنية ثابتة.

٤- الخطأ (Residual): يمثل ما تبقى من السلسلة وتعبير عن الأخطاء الحاصلة وتسمى بـ (White Noise). وهذه المركبة لا يمكننا استخدامها كدخل لعمليات التنبؤ، التحليل وبناء النماذج.

٧- الدراسات المرجعية:

اقترح الباحثون في [6] آلية للتنبؤ بحركة المرور بالاعتماد على فترات زمنية قصيرة باستخدام خوارزمية SVR، ولتحسين أداء هذه الخوارزمية المستخدمة تم استخدام الخوارزميات الجينية لضبط معاملاتها بهدف تحقيق أعلى دقة ممكنة، وتم تقسيم مجموعة البيانات المستخدمة لـ ٤٠ يوم من أيام العمل للتدريب و ١٠ للاختبار. وأظهرت النتائج أن استخدام الخوارزميات الجينية قد حسن نتائج التدريب مقارنة باستخدام خوارزمية SVR بشكل عادي وحقت معدل خطأ أقل من استخدام خوارزميات التحليل الإحصائي كـ ARIMA. واستخدام الباحثون في [7] بناء نظام يقوم بالتحكم بالإشارات المرورية بالاعتماد على التنبؤ بحركة المرور وقامت الدراسة بدراسة عدد من الخوارزميات كالانحدار التدريجي (Gradient boosting) والغابات العشوائية (random forest) وغيرها وأظهرت خوارزمية الغابات العشوائية أداءً أفضل من باقي الخوارزميات الأخرى، وساهمت هذا التطبيق في تخفيف الازدحام المروري بنسبة 30.8% من خلال محاكاة حركة طريق وإشارة مرورية عن طريق مكتبات تؤمنها لغة python. وقام الباحثون في [8] باستخدام مجموعة بيانات حركة المرور مع التلوث المقدمة من قبل مشروع citypulse لمدينة Aarhus في الدنمارك [17]، حيث تم معالجة البيانات ومن ثم استخدام نماذج مختلفة للتنبؤ بحركة المرور حيث تم دمج مجموعات البيانات بناء على عامل الزمن لاستخدامها للتنبؤ من خلال استخدام خوارزميات أقرب جوار وغيرها من الخوارزميات الأخرى للتنبؤ بحركة المرور وأظهرت خوارزمية أقرب جوار نتائج أفضل من غيرها من الخوارزميات، وقام الباحثون أيضاً ببناء نموذج معتمد على بناء مصنفات ضعيفة ومن ثم اختيار أفضل

N مصنف وتجميعها عن طريق خوارزميات التجميع لتحقيق تنبؤ أفضل؛ لم يتم تقييم أفضل قيمة لـ N في الدراسة السابقة. وتم دمج البيانات دون مراعاة عامل الزمن وعلاقة حركة المرور بالبيانات السابقة. أما الباحثون في [9] فاستخدموا مجموعة الخوارزميات للتنبؤ بحركة المرور لتحسين نظام التنقل ضمن المدن دون الحاجة لتجميع كم كبير من البيانات وأظهرت خوارزميات التعلم العميق نتائج أفضل على مستوى بعض المقاييس كـ R2 وساهمت الدراسة بشكل مشابه للدراسة [7] بتحسين حركة المرور ضمن المدينة واستخدمت خاصية عدد المركبات فقط دون دراسة خواص أخرى تؤثر على حركة المرور. قام الباحثون في [5] بعملية التنبؤ بقيم حركة المرور بالاعتماد على قيم الغازات المنبعثة (بيانات التلوث) ضمن المدينة الذكية حيث يتم استخدام القيم السابقة للحساسات في عملية التنبؤ ضمن مدينة Aarhus في الدنمارك [17]، أظهرت الدراسة أن خوارزميات التعلم العميق تحقق أداءً أفضل من باقي الخوارزميات المدروسة وأن حجم النافذة اللازم لتحقيق أفضل أداء للخوارزميات السابقة بلغ 132، ولم تستخدم الدراسات السابقة خواص أخرى تؤثر بشكل أكبر على التنبؤ كسرعة الحركة. ولم تدرس عملية التنبؤ بالاعتماد على القيم الحالية للحساسات فقط بالإضافة لعدم تحديد آلية ضبط البارامترات المستخدمة وآلية معالجة البيانات المفقودة.

٨- القسم العملي:

٨-1 مجموعة البيانات المستخدمة في الدراسة:

تم استخدام مجموعة البيانات من مشروع [17] citypulse EU FP7 لمدينة Aarhus في الدنمارك، حيث تم تثبيت مجموعة من الحساسات يبلغ عددها ١٣٦ حساس في بدايات ونهايات الطرق في المدينة بهدف تجميع بيانات تخص عدد المركبات التي تعبر الطريق، متوسط سرعتها بالإضافة لنسب بعض الغازات في الجو كـ (CO₂, NO₂, ...) خلال فواصل زمنية تبلغ ٥ دقائق، يتم تخزين البيانات ضمن ملفات بلاحقة CSV. لكل من حركة

المرور مع سرعة الحركة وملفات أخرى لنسب الغازات في الجو لكل طريق عن طريق استخدام بنية سحابية تربط الحساسات مع بنية التخزين [13]، وبلغ عدد التقارير المتولدة ٤٤٩ تقرير لكافة الطرق ضمن المدينة وذلك عن طريقة بنية لتجميع البيانات. تتوفر البيانات أيضاً بصيغة (tll). وهي عبارة ملفات البيانات السابقة موسمه دلاليماً باستخدام تقنيات الويب الدلالي [14, 15, 16] بصيغة (turtle) لبناء تطبيقات تساهم في تحسين تمثيل البيانات لتطوير تطبيقات المدن الذكية. يوضح الجدول [1] عينة من بيانات حركة المرور المسجلة باستخدام الحساسات.

الجدول (١) عينة من مجموعة بيانات حركة المرور المستخدمة [17]

REPORT_ID	avgSpeed	vehicleCount	TIMESTAMP
158324	50	5	2014-08-01T07:50:00
158324	50	6	2014-08-01T07:55:00
158324	60	4	2014-08-01T08:00:00
158324	60	1	2014-08-01T08:05:00

يشير حقل report id لمعرفة يخصص الطريق المدروس، كما ويوضح الجدول [2] عينة من بيانات التلوث المستخدمة والتي يتم جمع قراءتها أيضاً باستخدام الحساسات:

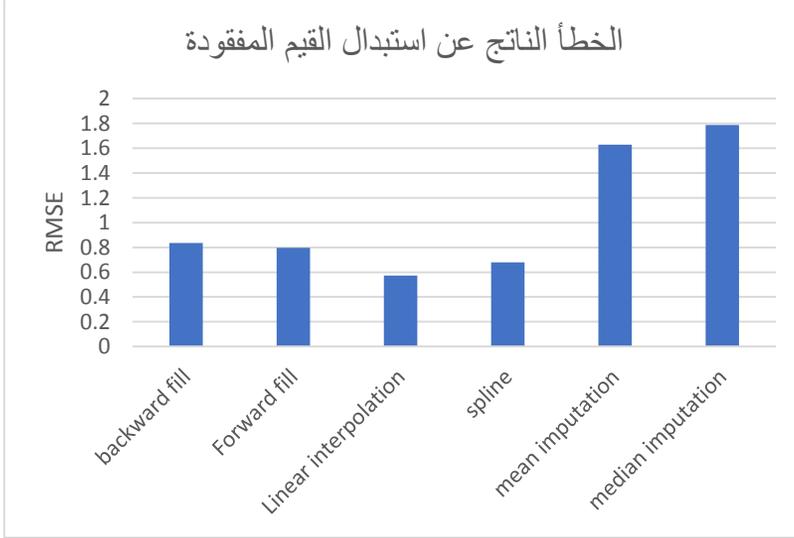
الجدول (٢) عينة من مجموعة البيانات التلوث المستخدمة [17]

latitude	longitude	NO ₂	SO ₂	CO ₂	PM	O ₃	TIMESTAMP
56.2317	10.10498	87	44	49	94	101	2014-08-01T07:50:00
56.2317	10.10498	86	47	48	97	106	2014-08-01T07:55:00
56.2317	10.10498	85	42	49	95	107	2014-08-01T08:00:00
56.2317	10.10498	87	44	51	90	103	2014-08-01T08:05:00

كما وتحوي مجموعة البيانات السابقة معلومات أخرى تخص الطقس، الأحداث في المدينة والمناسبات الاجتماعية والفعاليات ضمنها وحالة مواقف السيارات.

8-2 تجهيز البيانات Data Preparation:

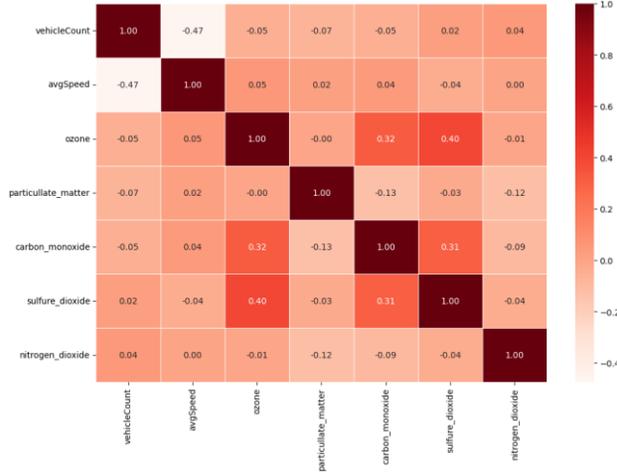
قبل تطبيق الخوارزميات السابقة على مجموعة البيانات لدينا سنقوم بتحميل البيانات ومعالجتها والتأكد من خلوها من المشاكل وذلك باستخدام مكتبة تحليل البيانات بلغة بايثون (Pandas) [18]. تعاني مجموعة البيانات السابقة من نقص في البيانات نتيجة عدم تمكن الحساسات من إرسال البيانات نتيجة مشكلة في الاتصال أو تم إهمال القراءة بسبب ضجيج البيانات. توجد العديد من الطرق لاستبدال القيم المفقودة ضمن مجموعة البيانات بشكل عام كاستبدال القيم المفقودة بقيمة ما معينة كالتوسط، أو الوسيط لكن مثل هذه الطرق لا تناسب طبيعة السلاسل الزمنية ويسبب تشوهاً للسلسلة خاصة إذا كانت تتمتع بخاصية الموسمية (الفصلية) [25]. سنقوم بمقارنة الطريقتين السابقتين مع طرق أخرى تستخدم لاستبدال القيم المفقودة ضمن السلاسل الزمنية كالأستبدال بأخر قيمة سابقة (Forward Fill)، أو الأستبدال بأخر قيمة لاحقة (Backward Fill) وهي مقدمة من قبل مكتبة pandas [18]، طريقة الإكمال الخطي (Linear Interpolation) [25] والتي تعتمد على وجود علاقة خطية بين القيم السابقة واللاحقة للقيمة المراد المفقودة المراد استبدالها، وأخيراً طريقة spline والتي تعتمد على وجود علاقة غير خطية بين القيم السابقة واللاحقة للقيمة المفقودة على عكس الطريقة السابقة. لتحديد الطريقة الأفضل قمنا بحذف بعض قيم البيانات الفعلية واستبدالها باستخدام إحدى الطرق السابقة، وبالأستفادة من البيانات الفعلية والبيانات الناتجة عن عملية الأستبدال قمنا بحساب الخطأ الناتج عن هذه العمليات كما يوضح الشكل [3]:



شكل (3) الخطأ الناتج عن استبدال القيم المفقودة للطرق السابقة.

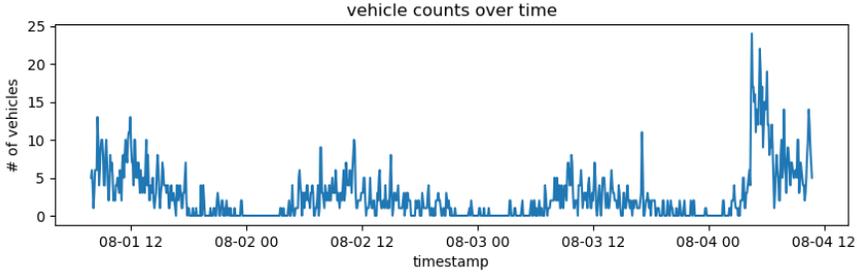
سنقوم بعد ذلك بعملية دمج البيانات مع بعضها البعض بالاعتماد على الوسمة الزمنية لتجهيز مجموعة البيانات لعملية التنبؤ، حيث سنستخدم بيانات التلوث وسرعة الحركة للتنبؤ بحركة المرور. يوضح الشكل [4] مصفوفة الترابط (correlation matrix) بين خواص الدخل والخرج المراد التنبؤ به.

تحليل حركة المرور ضمن المدن الذكية



الشكل (4) مصفوفة الترابط بين حركة المرور وخواص الدخل.

يُظهر الشكل [5] حركة المرور عبر الزمن لمدة ثلاثة أيام تقريباً حيث نلاحظ وجود موسمية في حركة البيانات حيث تكون حركة المرور مرتفعة خلال فترات النهار ومنخفضة نوعاً ما خلال فترات الليل.



الشكل (5) حركة المرور (عدد المركبات) عبر الزمن

3-8 الإعداد التجريبي:

الإعداد التجريبي الأول: وفيه نقوم بعملية التنبؤ بحركة المرور (عدد المركبات) بالاعتماد على القراءات الحالية للحساس خلال وسمة زمنية ما حيث تكون قيم القراءات في اللحظة

الحالية كدخل للخوارزمية وعدد المركبات هي القيمة المراد التنبؤ بها كما يوضح الجدول [3].

الجدول (3) البيانات المستخدمة للتنبؤ في التجربة الأولى

NO ₂	SO ₂	CO ₂	PM	O ₃	avgSpeed	vehicleCount (output)
87	44	49	94	101	50	5
86	47	48	97	106	50	6
85	42	49	95	107	60	4
87	44	51	90	103	60	1

الإعداد التجريبي الثاني: وفيه نقوم بالتنبؤ بحركة المرور بالاعتماد على القراءات الحالية للحساس بالإضافة للقراءات السابقة التي تم تسجيلها وتخزينها وتختلف عدد الخواص المدخلة هنا بحسب حجم النافذة المحدد، يوضح الجدول [4] شكل مجموعة البيانات من أجل قيمة النافذة تساوي ١ على سبيل المثال، بتعبير آخر بفرض أننا حصلنا على القراءات في لحظة ما (t) ونريد التنبؤ بحركة المرور في هذه اللحظة فإننا نستخدم قراءات الحساس المسجلة في هذه اللحظة واللحظة السابقة لها مباشرة (t-1).

الجدول (4) مثال عن البيانات المستخدمة للتنبؤ في التجربة الثانية

NO ₂	SO ₂	CO ₂	PM	O ₃	avgSpeed	NO ₂	SO ₂	CO ₂	PM	O ₃	avgSpeed	vehicleCount (output)
t-1	t-1	t-1	t-1	t-1	t-1	t	t	t	t	t	t	
٨٥	٤٢	٤٨	٩٢	١٠٠	٤٧	87	44	49	94	101	50	5
87	44	49	94	101	50	86	47	48	97	106	50	6
86	47	48	97	106	60	85	42	49	95	107	60	4
85	42	49	95	107	60	87	44	51	90	103	60	1

4-8 خوارزميات وعوامل التقييم المستخدمة في الدراسة:

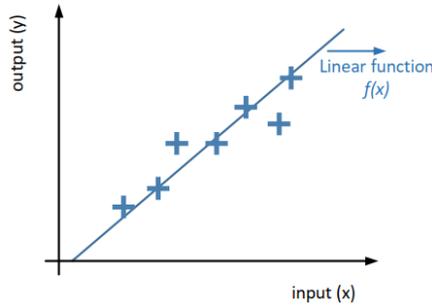
تستخدم خوارزميات الانحدار للتنبؤ بالقيم المستمرة (continuous values) وذلك من خلال إيجاد علاقة بين خواص الدخل والخرج. تدعم مكتبة تعلم الآلة sklearn (Science toolkit) [19] العديد من الخوارزميات التي يمكننا استخدامها نذكر منها:

1-4-8 الانحدار الخطي البسيط (Simple Linear Regression):

في هذا النوع من الخوارزميات تكون العلاقة التي تربط بين الدخل والخرج على شكل معادلة خط مستقيم كما في المعادلة [1]، حيث تكون قيمة (y) مثلاً هي حركة المرور التي نريد التنبؤ بها و (x) تعبر عن سرعة الحركة مثلاً.

$$y = mx + p \quad (1)$$

وتمثل قيم المتحولات (m, p) بارامترات التابع حيث تسعى خوارزمية الانحدار لإيجاد أفضل القيم المناسبة للبارامترين السابقين وذلك باستخدام خوارزميات أمثلية كالتدرج الاشتقاقي (gradient descent) بحيث تكون قيمة تابع الخسارة أقل ما يمكن. يوضح الشكل [6] مثلاً عن نموذج الانحدار الخطي [10].



الشكل (6) الانحدار الخطي البسيط [22]

8-4-2 الانحدار الخطي المتعدد (Multiple Linear Regression) [22]:

تستخدم هذه الخوارزمية في إيجاد العلاقة بين مجموعة من متحولات الدخل (x) ومتحول الخرج (y) المراد التنبؤ بقيمته، تعبر المعادلة [2] عن آلية حساب التنبؤ.

$$y = m_1x_1 + m_2x_2 + \dots + p \quad (2)$$

تساهم إضافة متحولات أخرى على عملية التنبؤ لتفسير الخرج بشكل أفضل مما يؤدي بشكل عام لتنبؤ أفضل من النوع السابق، لكن يجب أن تتم عملية إضافة المتحولات للمعادلة بشكل مدروس لتجنب الوصول لحالات مثل الملائمة الزائدة (Overfitting) وغيرها من المشاكل الأخرى.

8-4-3 خوارزمية ridge regression [24]:

تستخدم هذه الخوارزمية لحل مشاكل تعاني منها خوارزميات الانحدار التقليدية وتنتج هذه المشكلة عند إمكانية التنبؤ بقيمة متحول الخرج بعلاقة خطية عن طريق الدخل وبدقة عالية وتدعى بـ (multicollinearity)، عند حدوث المشكلة السابقة فإن النماذج المستخدمة قد تعطي نتائج غير دقيقة وبالتالي تملك هذه النماذج تباين عالي (high variance) فإن إضافة درجة من التباين للنموذج تؤدي لتحسين النتيجة، وهنا نعمل على إضافة قيمة معامل (α) بحيث يقلل من كلفة المركبات التي يتم إيجادها باستخدام خوارزمية الانحدار التقليدية. على سبيل المثال تكون معادلة تابع الخسارة والتي تمثل مجموع الفرق التربيعي بين القيم الفعلية والقيم المتنبئ بها وفق المعادلة [3]:

$$\widehat{\beta}_{ridge} = \arg \min \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \alpha \sum_{j=1}^p \beta_j^2 \right\} \quad (3)$$

حيث تعبر (n) عن عدد سجلات الدخل، (p) عدد الخواص المدروسة والمستخدم كدخل في عملية التنبؤ، (β_0, β_j) المركبات الناتجة أثناء تدريب الخوارزمية.

8-4-4 خوارزمية الشبكات المرنة (Elastic Nets) [24]:

تستفيد هذه الخوارزمية من الخوارزمية السابقة وخوارزمية lasso regression والتي تساعد في تحديد الخواص الأكثر تأثيراً على الخرج بدمج مزايا كل منها من خلال احتوائها على المعامل (α) الذي يساهم في ضبط كلفة المركبات الناتجة عن الخوارزمية والمعامل (β) الذي يعرف بـ (L1 ratio) والذي يحدد لنا الخواص الملائمة للاستخدام في عملية التنبؤ دوناً عن غيرها وتعطى المعادلة [3] وفق الشكل التالي:

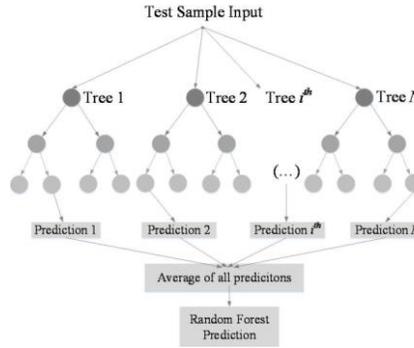
$$\widehat{\beta}_{EN} = \arg \min \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \alpha (\beta \sum_{j=1}^p |\beta_j| + (1 - \beta) \sum_{j=1}^p \beta_j^2) \right\} \quad (4)$$

8-4-5 خوارزمية أشجار القرار (Decision Trees) [22]:

تستخدم هذه الخوارزمية أيضاً للتنبؤ بالقيم الفئوية والمستمرة، وتحتوي الشجرة على مجموعة من العقد الداخلية التي تمثل قيم خاصة ما يتم إجراء الاختبار عليها لتحديد الفرع الذي سنسلكه ضمن الشجرة، بالإضافة لعقد أوراق تحوي القيمة الرقمية التي تنبئت بها الشجرة. يتم اختيار العقدة عن طريق معايير معينة نحددها عن طريق بارامتر (splitter) عند استخدام تحقيق الخوارزمية وعادة ما نستخدم القيمة (best) بدلاً من (random) لتحقيق أفضل طريقة لتقسيم الشجرة. تعمل الخوارزمية وفق آلية تحقق من قيمة الدخل والمروور على فروع الشجرة بحيث يكون لدينا معرفة أكبر عن الدخل لتحديد القيمة المناسبة للتنبؤ. توجد العديد من البارامترات التي يمكن ضبطها من أجل شجرة القرار أهمها ارتفاع الشجرة (max depth)، عدد العينات اللازمة لتقسيم الأوراق (min_sample_split) وأخيراً قيمة (α) والذي يستخدم لإعادة الشجرة الأقل كلفة لتجنب الملائمة الزائدة بحيث يتم اختيار الشجرة التي تملك أكبر كلفة وأقل من قيمة (α).

8-4-6 الغابات العشوائية (Random Forest) [22]:

يتم بناء هذه الغابة بالاعتماد على عدد من أشجار القرار التي تستخدم في عملية التنبؤ ويتم ضبطها عن طريق بارامتر ($n_estimators$) ويمكننا تحديد عمق كل شجرة بشكل مشابه لخوارزمية أشجار القرار. ولحساب التنبؤ النهائي يتم أخذ متوسط القيم التي يتم التنبؤ بها من قبل كل شجرة ضمن هذه الغابة. تساهم هذه الخوارزمية في تجنب مشكلة الملائمة الزائدة من خلال بناء أشجار بسيطة مع اعتماد هذه الأشجار على خواص أقل، كما ويمكن أن تعاني من مشاكل عند وجود ضجيج في البيانات كما يوضح الشكل [7]:

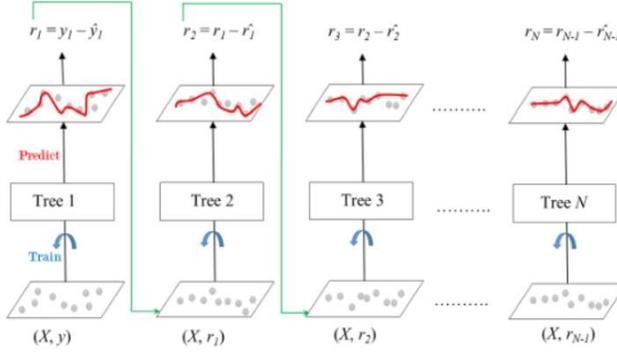


الشكل (7) التنبؤ بالقيم المستمرة باستخدام الغابات العشوائية [22]

8-4-7 خوارزمية الانحدار التدريجي (Gradient Boosting Regression):

تبدأ هذه الخوارزمية بإجراء تنبؤ أولي بالاعتماد على جميع بيانات الدخل ويساوي متوسط قيمة متحول الخرج لكل من البيانات المدخلة للخوارزمية، والخطوة التالية هي بناء شجرة من خلال دراسة تحليل الدخل وتحديد الخواص التي سيتم بناء الشجرة على أساسها بهدف تقليل قيمة تابع الخسارة ($loss$ function) الذي يمثل الفرق بين القيم الفعلية والقيم المتنبئ بها في كل مرحلة. يتم جمع الأشجار الناتجة في كل مرحلة لبناء المصنف النهائي بحيث

لم يعد بإمكان المصنف الناتج تحسين قيمة التنبؤ الناتجة [22]. يوضح الشكل [8] آلية عمل خوارزمية الانحدار التدريجي.

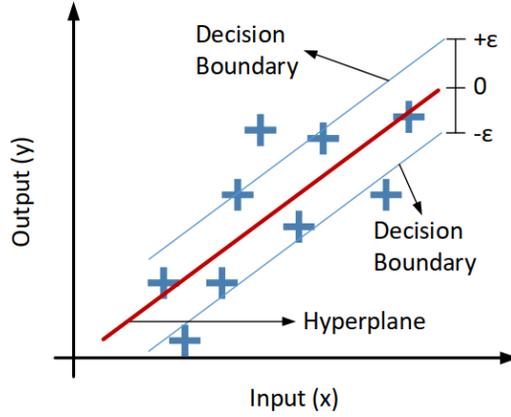


الشكل (8) التنبؤ بالقيم المستمرة باستخدام الانحدار التدريجي [23]

يتم ضبط هذه الخوارزمية بشكل مشابه لخوارزمية الغابات العشوائية، وتحتوي أيضاً على بعض البارامترات الأخرى كمعدل التعلم (learning rate) والذي يقوم بضبط تأثير الخوارزميات الضعيفة المبنية عليها الخوارزمية على النتيجة النهائية.

8-4-8 خوارزمية support vector regression [22]:

تشتهر هذه الخوارزمية باستخدامها ضمن مسائل التصنيف (classification) ولكنها تستخدم في بعض الأحيان كخوارزمية للتنبؤ بالقيم المستمرة، وتعتمد هذه الخوارزمية على مجموعة من البارامترات المهمة كالنواة (kernel) والتي تستخدم لتحديد السطح (hyperplane) مع تحقيق كلفة حساب منخفضة، فيما يستخدم السطح للفصل بين القيم أما البارامتر الثالث فهو حدود القرار (decision boundary) وهو عبارة عن خطين متوازيين واحد للقيم الموجبة والآخر للقيم السالبة كما يوضح الشكل [9].



الشكل (9) التنبؤ بالقيم المستمرة باستخدام SVR [22]

تبحث الخوارزمية عن أفضل سطح يحوي أكبر عدد من النقاط بحيث يكون هو الخط المستخدم في عملية التنبؤ على عكس الخوارزميات الأخرى التي تبحث عن أقل معدل خطأ بين القيم الفعلية (actual values) والقيم الناتجة عن التنبؤ (predicted values)، بحيث لا يتجاوز الخط المراد البحث عنه قيمة العتبة (ϵ) وتعد هذه الخوارزمية سهلة التحقيق ودقة عالية في التنبؤ لكنها تعاني مع حجم البيانات الكبير أو عند وجود ضجيج في البيانات [22].

لتقييم أداء الخوارزميات السابقة سنستخدم بعض المعايير كـ (Root Mean Squared Error) (RMSE) وهو عبارة عن الجذر التربيعي لمتوسط الفروق بين القيم الحقيقية (O_i) والقيم الناتجة عن التنبؤ (P_i) ويعطى بالمعادلة [5]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (5)$$

يمثل المعيار السابق الجذر التربيعي لمربع الفروق بين القيم الفعلية والقيمة الناتجة عن التنبؤ وبالتالي يوضح مدى قرب القيم الناتجة عن التنبؤ من القيم الفعلية. كما ويمكن

استخدام معامل Mean Absolute Error (MAE) والذي يعبر عن المتوسط الحسابي للقيمة المطلقة للفروق بين القيم الحقيقية والقيم الناتجة عن التنبؤ.

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i|^2 \quad (6)$$

أخيراً يمكن استخدام معامل R2 ويعطى بالمعادلة التالية:

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}, \text{ where } \bar{O} = \frac{1}{n} \sum_{i=1}^n O_i \quad (7)$$

يسمى المعيار السابق بمركبة التحديد (Coefficient of Determination) ويوضح المعامل السابق مدى مناسبة البيانات لنموذج الانحدار أو مدى جودة تفسير البيانات الموجودة لدينا باستخدام متحول الدخل وكلما كانت القيمة أكبر يدل على أداء أفضل للخوارزمية على عكس المعاملين السابقين.

كما ذكرنا سابقاً تمتلك هذه الخوارزميات مجموعة من البارامترات التي يجب علينا ضبطها لأجل التحكم بالنموذج الناتج [7]، وقمنا في هذه الدراسة بتعديل هذه القيم عن طريق البحث عن أفضل قيم للبارامترات السابقة ضمن مجموعة القيم المحتملة لكل منها بهدف الحصول على أقل معدل خطأ تنبؤ (RMSE). يوضح الجدول [5] القيم التي تم الحصول عليها عن طريق البحث عن أفضل القيم الممكنة لكل من بارامترات الخوارزميات السابقة.

الجدول (5) البارامترات المستخدمة مع الخوارزميات المدروسة

المعاملات	الخوارزمية
alpha=0.005	Ridge Regression
alpha=0.005, l1_ratio=0	ElasticNet Regression
random_state=42, max_depth=3 min_samples_split=2 ccp_alpha = 0, splitter = 'best'	Decision Tree

random_state=42, max_depth=10 n_estimators=20 min_samples_split=5	Random Forest
random_state=42, max_depth=2 n_estimators=20, min_samples_split=2 learning_rate = 0.3	Gradient Boosting
kernel='rbf', C=1	SVR

8-5 تأثير حجم النافذة (Window Size) على عملية التنبؤ:

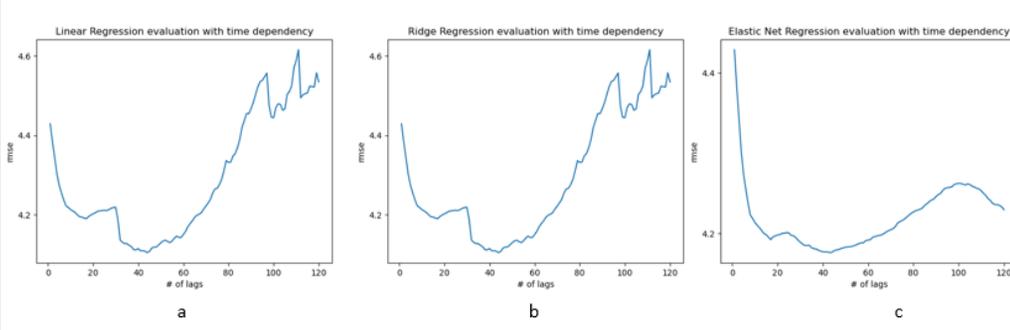
سنقوم الآن بدراسة تأثير حجم النافذة والتي تعبر عن عدد قراءات الحساسات السابقة التي يجب استخدامها بحيث نحقق أفضل نتيجة تنبؤ ممكنة. وأظهرت التجربة أن إضافة سرعة حركة المرور مع بيانات التلوث تساهم في تقليل حجم النافذة اللازمة للتنبؤ بقيم حركة المرور عن الدراسة [5] ويحسن أداء النماذج بشكل عام ويوضح الجدول [6] نتائج اختبار الخوارزميات مع حجم النافذة الأفضل لكل منها.

الجدول (6) أفضل نافذة ممكنة من أجل الخوارزميات المدروسة.

الخوارزمية	أفضل قيمة للنافذة	قيمة الخطأ (RMSE)
Linear Regression (LR)	44	4.104438
Ridge Regression	٤٤	٤,١٠٤٤٣٨
ElasticNet Regression (EN)	43	4.175865
Decision Tree	99	٤,٢٧٧٦٥٩
Gradient Boosting (GB)	104	3.827373
Random Forest (RF)	115	3.630251
SVR	104	٣,٨٢٧٣٧٣

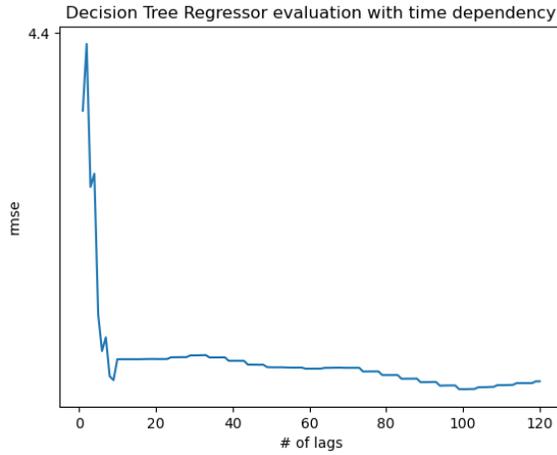
تحليل حركة المرور ضمن المدن الذكية

يوضح الشكل [10] تأثير حجم النافذة على عملية التنبؤ بحركة المرور حيث أنه بعد القيمة ٤٤ تزداد قيمة الخطأ المرتكب في عملية التنبؤ وذلك من أجل الخوارزميات (linear regression, ridge, elasticNets).

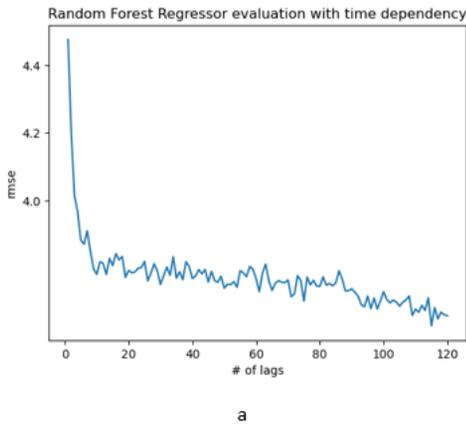


الشكل [10] تأثير حجم النافذة على خوارزميات التنبؤ: (a): LR, (b): ridge, (c):(EN)

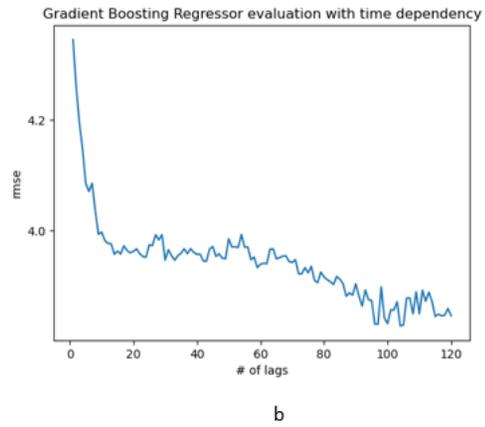
أما خوارزمية شجرة القرار فتحتاج لبيانات أكثر حتى تعطي نتيجة أكثر دقة حيث بينت الدراسة أن هذه الخوارزمية تقدم أداءً أسوأ من الخوارزميات السابقة ويوضح الشكل [11] تأثير حجم النافذة على عملية التنبؤ من أجل خوارزمية شجرة القرار. كما ونلاحظ من الشكل [12] أن خوارزميات التجميع تساهم في تحسين عملية التنبؤ بشكل ملحوظ حيث تحتاج لحجم نافذة أقل من باقي الخوارزميات المدروسة من جهة ومن جهة أخرى تحقق أقل معدل خطأ مقارنة بالخوارزميات الأخرى المستخدمة نظراً لفعالية آلية التنبؤ المستخدمة ضمن هذه الخوارزميات.



الشكل [11] تأثير حجم النافذة على خوارزمية أشجار القرار



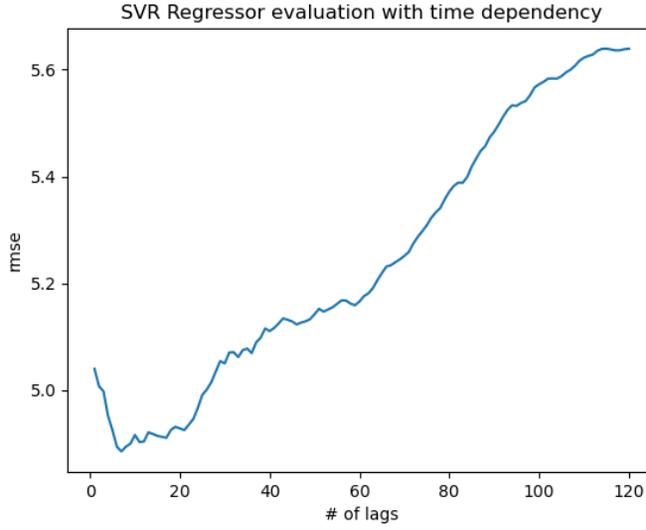
a



b

الشكل [12] تأثير حجم النافذة على خوارزميات التنبؤ: (a): RF, (b): GB

ويوضح الشكل [13] أخيراً أداء خوارزمية SVR والتي تحقق أداءً أسوأ من باقي الخوارزميات وتحقق أداءً جيداً من أجل حجم نافذة أقل لكن حجم الخطأ أكبر من باقي الخوارزميات.



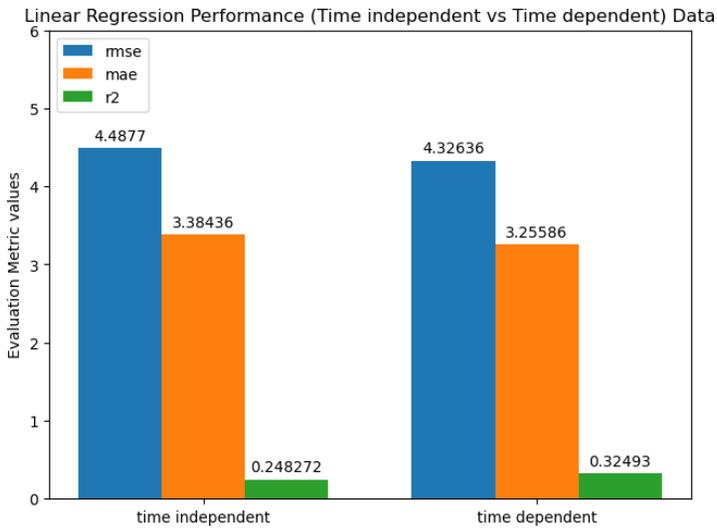
الشكل [13] تأثير حجم النافذة على خوارزمية SVR.

8-6 تأثير استخدام القراءات السابقة في عملية التنبؤ:

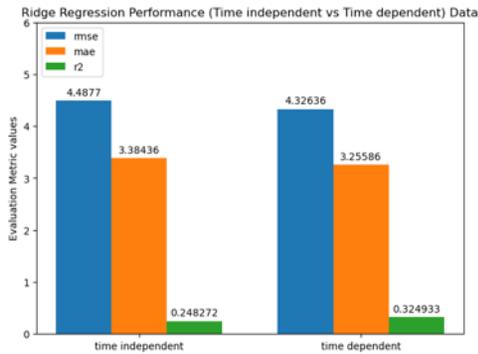
تحتاج معظم الخوارزميات لحجم نافذة 44 على الأقل حتى نبدأ بالحصول على أقل معدل خطأ ممكن ويتحسن أداء عملية التنبؤ، لذا ولمقارنة تأثير استخدام القراءات السابقة سنقوم بحساب متوسط الخطأ الناتج عن استخدام القراءات السابقة للحساسات من حجم النافذة 44 وحتى 120 ومقارنة نتائج الخطأ مع استخدام القراءات الحالية للحساسات.

يظهر الشكل [14] الخطأ الناتج من أجل عوامل التقييم المدروسة لأجل خوارزمية Linear Regression حيث يظهر استخدام قراءات الحساس السابقة تحسن علمية التنبؤ بشكل بسيط، كما تشير قيمة r^2 score لأن النموذج الناتج عن استخدام القراءات السابقة أفضل.

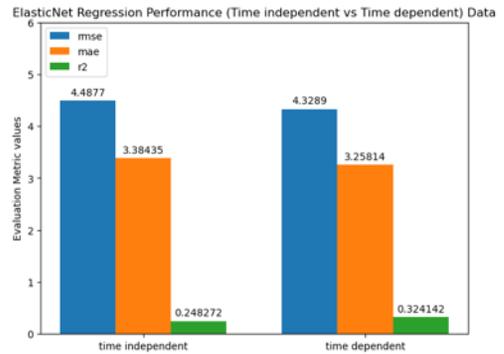
ويظهر الشكل [15] الخطأ الناتج من أجل خوارزميتي ridge و ElasticNets حيث يتشابه أداء الخوارزميات السابقة عند استخدام القراءات الحالية للحساس فقط. فيما تحقق خوارزمية ridge أداءً أفضل عند استخدام قراءات الحساس السابقة في التنبؤ.



الشكل [14] تقييم أداء خوارزمية (LR).



a

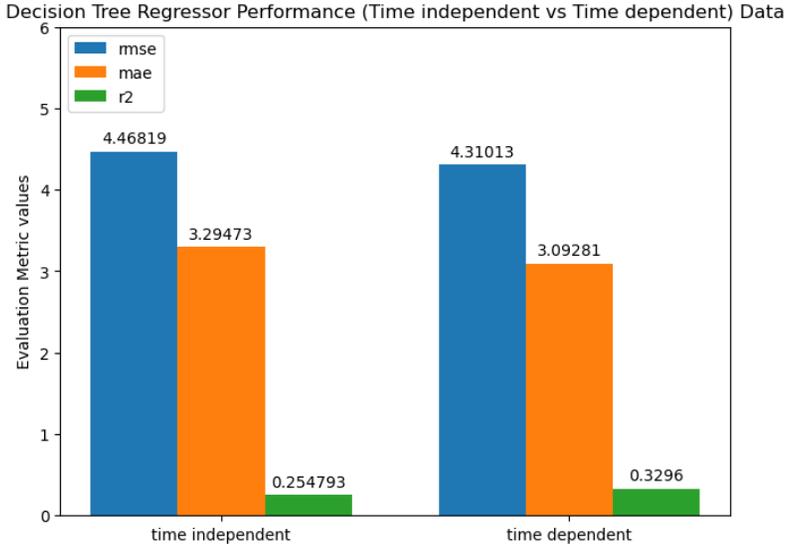


b

الشكل [15] تقييم أداء خوارزميتي: EN: (a): ridge, (b): EN

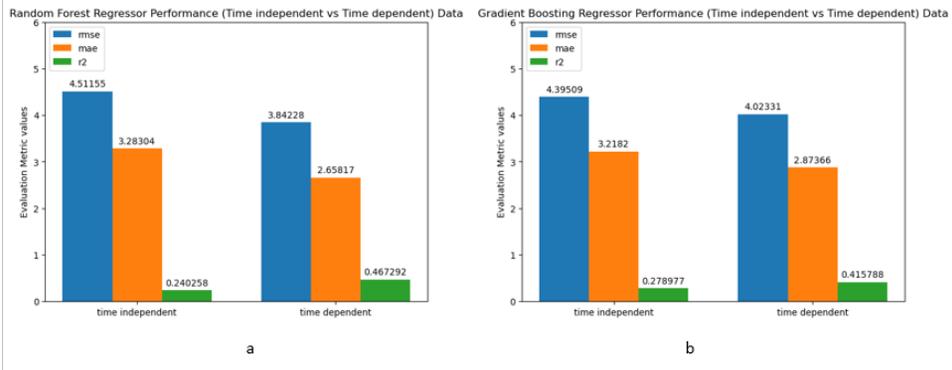
تحليل حركة المرور ضمن المدن الذكية

فيما يظهر الشكل [16] الخطأ الناتج عن استخدام خوارزمية شجرة القرار والتي يتحسن أدائها عند استخدام القراءات السابقة للحساس، فيما تقدم أداءً أسوأ من الخوارزميات السابقة من أجل التجريبتين المدروستين.



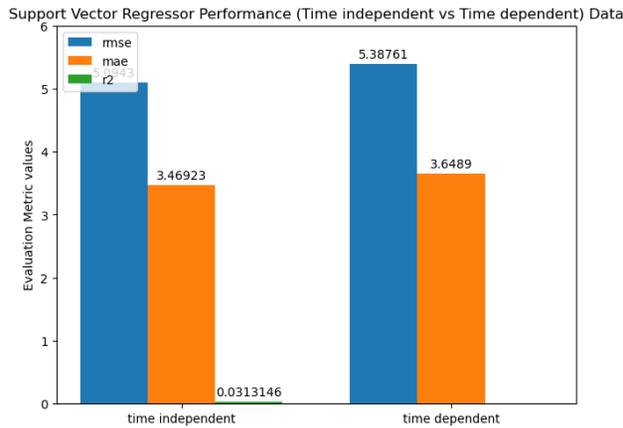
الشكل [16] تقييم أداء خوارزمية (DT).

يظهر الشكل [17] تقييم المعاملات السابقة من أجل خوارزميتي (RF, GB) حيث يتحسن أداء خوارزمية RF بشكل كبير عند استخدام بيانات أكبر من استخدام القراءات الحالية للحساس فقط، فيما يتحسن أداء خوارزمية GB كباقي الخوارزميات لكنها لا تقدم أداءً أفضل من خوارزمية RF بشكل عام.



الشكل [17] تقييم أداء خوارزميتي: (a): RF, (b): GB

أخيراً يظهر الشكل [18] تقييم أداء خوارزمية SVR والتي تظهر لنا أنها تحقق أداءً أسوأ من باقي الخوارزميات بشكل ملحوظ.



الشكل [18] تقييم أداء خوارزمية (DT).

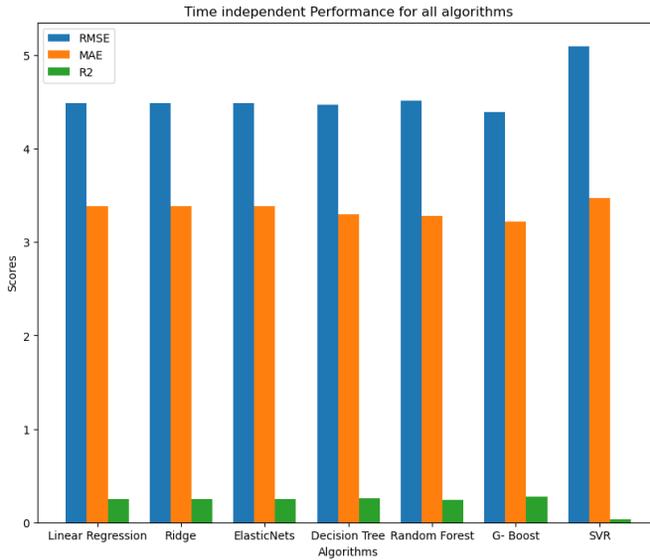
يوضح الجدول [7] ناتج تقييم أداء الخوارزميات السابقة بعد تنفيذ التجريبتين المدروستين:

تحليل حركة المرور ضمن المدن الذكية

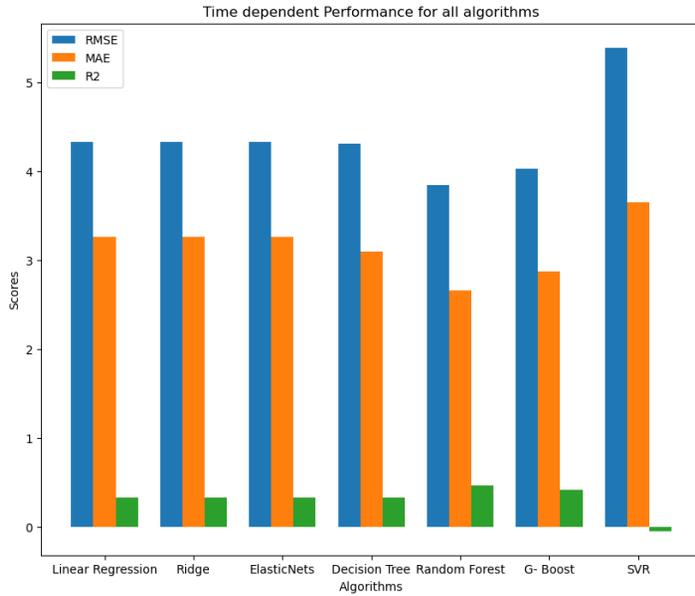
الجدول (7) تقييم أداء الخوارزميات المدروسة.

التجربة الثانية			التجربة الأولى			الخوارزمية
R2	MAE	RMSE	R2	MAE	RMSE	
0.3249	3.2558	4.3264	0.24827	3.3844	4.4877	LR
0.3249	3.2558	4.3264	0.24827	3.3844	4.4877	Ridge
0.32414	3.2582	4.3289	0.24827	3.3844	4.4877	EN
0.3296	3.0928	4.3101	0.25479	3.2947	4.4682	DT
0.46729	2.6582	3.8423	0.24025	3.283	4.5115	RF
0.41578	2.8736	4.0233	0.27897	3.2182	4.3951	GB
-0.001	3.6489	5.3876	0.03131	3.4692	5.0943	SVR

يوضح كل من الشكلين [19] و [20] نتائج مقارنة أداء الخوارزميات السابقة من أجل التجريبتين المدروستين.

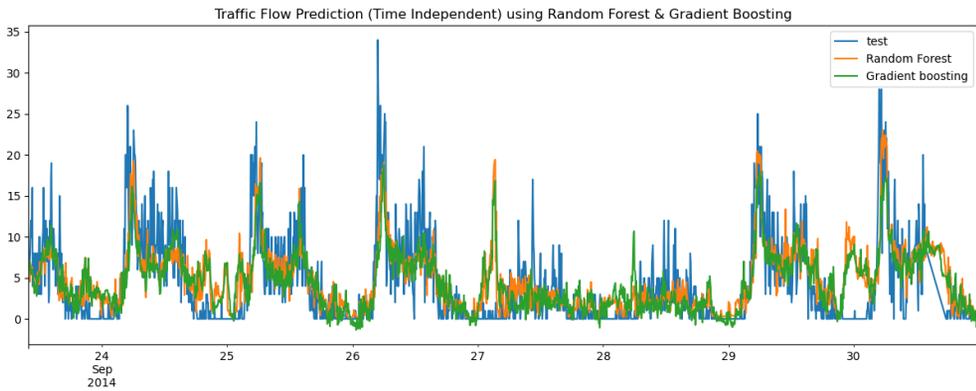


الشكل [19] تقييم أداء الخوارزميات المدروسة عند استخدام القراءات الحالية للحساس.



الشكل [20] تقييم أداء الخوارزميات المدروسة عند استخدام القراءات الحالية والسابقة للحساس.

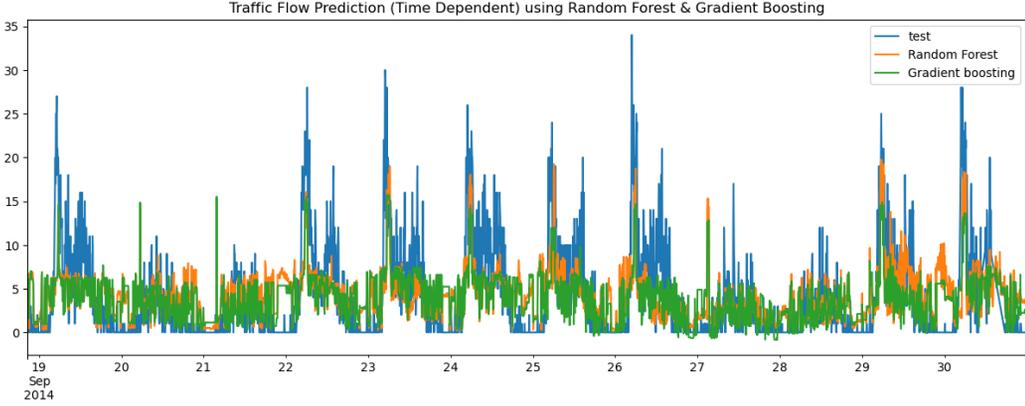
يوضح الشكل [21] السلسلة الزمنية الناتجة عن التنبؤ باستخدام كل من خوارزميتي الغابات العشوائية والانحدار التدريجي مع بيانات الاختبار دون الأخذ بعين الاعتبار القراءات السابقة للحساس في عملية التنبؤ.



الشكل [21] السلاسل الناتجة عن عملية التنبؤ بدون الاعتماد على قراءات الحساس السابقة

تحليل حركة المرور ضمن المدن الذكية

ويوضح الشكل [22] السلسلة الزمنية الناتجة عن التنبؤ باستخدام الخوارزميتين السابقتين مع بيانات الاختبار مع الأخذ بعين الاعتبار القراءات السابقة للحساس.



الشكل [22] السلاسل الناتجة عن عملية التنبؤ بالاعتماد على قراءات الحساس السابقة

٩- مناقشة النتائج:

أظهرت التجارب السابقة أن خوارزميات التجميع تحقق أداءً أفضل من باقي الخوارزميات، نظراً لطريقتها المتبعة في تحديد القيمة المراد التنبؤ بها حيث تعتمد على متوسطات القيم الناتجة عن المصنفات المستخدمة ضمناً أو كما في خوارزمية GB التي تحاول إيجاد أفضل نموذج ممكن يحقق أقل كلفة ممكنة. كما وتقدم الخوارزميات البسيطة أداءً جيداً مقارنة بخوارزمية SVR التي تقدم أداءً سيئاً مقارنة بباقي الخوارزميات نظراً لحجم البيانات الكبير المدروس.

١٠- المقترحات والتوصيات المستقبلية:

قدمت الدراسة السابقة تقيماً لأداء خوارزميات regression في تحليل حركة المرور وأظهرت أن هناك خوارزميات يمكن أن تحقق نتائج في عملية التنبؤ أفضل من غيرها، يمكن أن يتم تطوير هذه الدراسة باستخدام خوارزميات التعلم العميق ومقارنتها مع

الخوارزميات السابقة بهدف الحصول على تنبؤ أفضل بالإضافة لإمكانية دراسة تأثير حركة المرور الأخرى في الطرق المجاورة بهدف معرفة تأثيرها على حركة الطريق المدروس.

- [1] Road traffic injuries report, World Health Organization, <https://www.who.int/en/news-room/fact-sheets/detail/road-traffic-injuries> last access 7/12/2024
- [2] Tsuboi, T., 2021. Traffic Flow Analysis and Management. In Design of Cities and Buildings–Sustainability and Resilience in the Built Environment. IntechOpen.
- [3] Muthuramalingam, S., Bharathi, A., Rakesh Kumar, S., Gayathri, N., Sathiyaraj, R. and Balamurugan, B., 2019. IoT based intelligent transportation system (IoT-ITS) for global perspective: A case study. Internet of things and big data analytics for smart generation, pp.279–300.
- [4] Yuan, H. and Li, G., 2021. A survey of traffic prediction: from spatio-temporal data to intelligent transportation. Data Science and Engineering, 6(1), pp.63–85.
- [5] Shouaib, M., Metwally, K. and Badran, K., 2023. An Enhanced Time-dependent Traffic Flow Prediction in Smart Cities. Advances in Electrical & Computer Engineering, 23(3).
- [6] Zhan, A., Du, F., Chen, Z., Yin, G., Wang, M. and Zhang, Y., 2022. A traffic flow forecasting method based on the GA-SVR. Journal of High Speed Networks, 28(2), pp.97–106.
- [7] Moumen, I., Abouchabaka, J. and Rafalia, N., 2023. Adaptive traffic lights based on traffic flow prediction using machine learning models.

International Journal of Electrical and Computer Engineering (IJECE),
13(5), pp.5813–5823.

[8] Khan, N.U., Shah, M.A., Maple, C., Ahmed, E. and Asghar, N., 2022. Traffic flow prediction: an intelligent scheme for forecasting traffic flow using air pollution data in smart cities with bagging ensemble. Sustainability, 14(7), p.4164.

[9] Navarro–Espinoza, A., López–Bonilla, O.R., García–Guerrero, E.E., Tlelo–Cuautle, E., López–Mancilla, D., Hernández–Mejía, C. and Inzunza–González, E., 2022. Traffic flow prediction for smart traffic lights using machine learning algorithms. Technologies, 10(1), p.5.

[10] Saleh, H. and Layous, J.A., 2022. Machine Learning–Regression (Doctoral dissertation, PhD thesis, Jan. 2022. doi: 10.13140/RG. 2.2. 35768.67842).

[11] Jiawei, H., Jian, P. and Hanghang, T., 2023. Data mining: concepts and techniques. Morgan kaufmann.

[12] Piri, M., Review of Regression Algorithms.

[13] Tönjes, R., Barnaghi, P., Ali, M., Mileo, A., Hauswirth, M., Ganz, F., Ganea, S., Kjærgaard, B., Kuemper, D., Nechifor, S. and Puiu, D., 2014, June. Real time iot stream processing and large–scale data analytics for smart city applications. In poster session, European Conference on Networks and Communications (p. 10). sn.

- [14] Bischof, S., Karapantelakis, A., Nechifor, C.S., Sheth, A.P., Mileo, A. and Barnaghi, P., 2014. Semantic modelling of smart city data.
- [15] Kolozali, S., Bermudez-Edo, M., Puschmann, D., Ganz, F. and Barnaghi, P., 2014, September. A knowledge-based approach for real-time iot data stream annotation and processing. In 2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) (pp. 215-222). IEEE.
- [16] Arenas, M., Heflin, J., Groth, P., Thirunarayan, K., d'Aquin, M., Strohmaier, M., Corcho, O., Staab, S., Srinivas, K., Simperl, E. and Dumontier, M., 2015. The Semantic Web-ISWC 2015. Springer International Publishing.
- [17] CityPulse Dataset Collection · A collection of semantically annotated datasets for the CityPulse EU FP7 Project <http://iot.ee.surrey.ac.uk:8080/index.html> last access 12/7/2024
- [18] The pandas development team (2024) "pandas-dev/pandas: Pandas". Zenodo. doi: 10.5281/zenodo.10957263.
- [19] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, pp.2825-2830.

- [20] J. D. Hunter, "Matplotlib: A 2D Graphics Environment", *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007
- [21] Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. *Nature* 585, 357–362 (2020).
- [22] Kinaneva, D., Hristov, G., Kyuchukov, P., Georgiev, G., Zahariev, P. and Daskalov, R., 2021, June. Machine learning algorithms for regression analysis and predictions of numerical data. In *2021 3rd International Congress on Human–Computer Interaction, Optimization and Robotic Applications (HORA)* (pp. 1–6). IEEE.
- [23] Rozas–Rodriguez, W., Pastor–Vargas, R., Peacock, A.D., Kane, D. and Carpio–Ibañez, J., 2024. BESS Reserve Optimisation in Energy Communities. *Sustainability*, 16(18), p.8017.
- [24] Nur, A.R., Jaya, A.K. and Siswanto, S., 2024. Comparative Analysis of Ridge, LASSO, and Elastic Net Regularization Approaches in Handling Multicollinearity for Infant Mortality Data in South Sulawesi. *Jurnal Matematika, Statistika dan Komputasi*, 20(2), pp.311–319.
- [25] Darji, J., Biswas, N., Jones, L.D. and Ashili, S., 2023. Handling missing data in the time–series data from wearables. In *Time Series Analysis–Recent Advances, New Perspectives and Applications*. IntechOpen.
- [26] Peixeiro, M., 2022. Time series forecasting in python. Simon and Schuster.

[27] م. حلا إبراهيم، د. رانيا لطفي، د. أسماء شعار، ٢٠٢٣. تحليل السلاسل الزمنية لاكتشاف الشذوذ ضمن حركة المرور. مجلة جامعة البعث-سلسلة العلوم الهندسية الميكانيكية والكهربائية والمعلوماتية، ٤٥ (٢٠).

[28] م. آلاء السباعي، د. محسن حسين، د. وسيم رمضان، ٢٠٢٢. أنظمة تحليل بيانات الزمن الحقيقي الضخمة الناتجة عن انترنت الأشياء. مجلة جامعة البعث-سلسلة العلوم الهندسية الميكانيكية والكهربائية والمعلوماتية، ٤٤ (١٤).