مجلة جامعة حمص

سلسلة العلوم الميكانيكية والكهربائية والمعلوماتية

مجلة علمية محكمة دورية

المجلد 47 . العدد ٢

1447 هـ - 2025 م

الأستاذ الدكتور **طارق حسام الدين** رئيس جامعة حمص المدير المسؤول عن المجلة

أ. د. وليد حمادة	رئيس تحرير مجلة جامعة حمص للعلوم الإنسانية
أ. د. درغام سلوم	رئيس تحرير مجلة جامعة حمص للعلوم الطبية والهندسية والأساسية والتطبيقية

د.محمد فراس رمضان	عضو هيئة التحرير
د. مضر سعود	عضو هيئة التحرير
د. ممدوح عبارة	عضو هيئة التحرير
د. موفق تلاوي	عضو هيئة التحرير
د. طلال رزوق	عضو هيئة التحرير
د. أحمد الجاعور	عضو هيئة التحرير
د. الياس خلف	عضو هيئة التحرير
د. روعة الفقس	عضو هيئة التحرير
د. محمد الجاسم	عضو هيئة التحرير
د. خليل الحسن	عضو هيئة التحرير
د. هيثم حسن	عضو هيئة التحرير
د. أحمد حاج موسى	عضو هيئة التحرير

تهدف المجلة إلى نشر البحوث العلمية الأصيلة، ويمكن للراغبين في طلبها الاتصال بالعنوان التالي:

رئيس تحرير مجلة جامعة حمص

سورية . حمص . جامعة حمص . الإدارة المركزية . ص . ب (٧٧)

. هاتف / فاكس : ٢١٣٨٠٧١ ٣١ ٩٦٣ ++

. موقع الإنترنت : www.homs-univ.edu.sy

. البريد الالكتروني : journal.homs-univ.edu.sy

ISSN: 1022-467X

شروط النشر في مجلة جامعة البعث

الأوراق المطلوبة:

- ٢ نسخة ورقية من البحث بدون اسم الباحث / الكلية / الجامعة) + CD / word من البحث منسق حسب شروط المجلة.
- طابع بحث علمي + طابع نقابة معلمين.
- **اذا كان الباحث طالب دراسات عليا:**

يجب إرفاق قرار تسجيل الدكتوراه / ماجستير + كتاب من الدكتور المشرف بموافقته على النشر في المجلة.

- **اذا كان الباحث عضو هيئة تدريسية:**

يجب إرفاق قرار المجلس المختص بإنجاز البحث أو قرار قسم بالموافقة على اعتماده حسب الحال.

- **اذا كان الباحث عضو هيئة تدريسية من خارج جامعة البعث :**

يجب إحضار كتاب من عمادة كليته تثبت أنه عضو بالهيئة التدريسية و على رأس عمله حتى تاريخه.

- **اذا كان الباحث عضواً في الهيئة الفنية :**

يجب إرفاق كتاب يحدد فيه مكان و زمان إجراء البحث ، وما يثبت صفته وأنه على رأس عمله.

- **يتم ترتيب البحث على النحو الآتي بالنسبة لكليات (العلوم الطبية والهندسية والأساسية والتطبيقية):**

عنوان البحث . . ملخص عربي و إنكليزي (كلمات مفتاحية في نهاية الملخصين).

١- مقدمة

٢- هدف البحث

٣- مواد وطرق البحث

٤- النتائج ومناقشتها .

٥- الاستنتاجات والتوصيات .

٦- المراجع.

- **يتم ترتيب البحث على النحو الآتي بالنسبة لكليات (الآداب – الاقتصاد– التربية – الحقوق – السياحة – التربية الموسيقية وجميع العلوم الإنسانية):**
- عنوان البحث . . ملخص عربي و إنكليزي (كلمات مفتاحية في نهاية الملخصين).

١. مقدمة.
٢. مشكلة البحث وأهميته والجديد فيه.
٣. أهداف البحث و أسئلته.
٤. فرضيات البحث و حدوده.
٥. مصطلحات البحث و تعريفاته الإجرائية.
٦. الإطار النظري و الدراسات السابقة.
٧. منهج البحث و إجراءاته.
٨. عرض البحث و المناقشة والتحليل
٩. نتائج البحث.
١٠. مقترحات البحث إن وجدت.
١١. قائمة المصادر والمراجع.

٧- يجب اعتماد الإعدادات الآتية أثناء طباعة البحث على الكمبيوتر:

أ- قياس الورق ١٧,٥×٢٥ B5.

ب- هوامش الصفحة: أعلى ٢,٥٤- أسفل ٢,٥٤ – يمين ٢,٥- يسار ٢,٥ سم

ت- رأس الصفحة ١,٦ / تذييل الصفحة ١,٨

ث- نوع الخط وقياسه: العنوان . Monotype Koufi قياس ٢٠

ـ كتابة النص Simplified Arabic قياس ١٣ عادي ـ العناوين الفرعية Simplified Arabic قياس ١٣ عريض.

ج . يجب مراعاة أن يكون قياس الصور والجداول المدرجة في البحث لا يتعدى ١٢سم.

٨- في حال عدم إجراء البحث وفقاً لما ورد أعلاه من إشارات فإن البحث سيهمل ولا يرد البحث إلى صاحبه.

٩- تقديم أي بحث للنشر في المجلة يدل ضمناً على عدم نشره في أي مكان آخر، وفي حال قبول البحث للنشر في مجلة جامعة البعث يجب عدم نشره في أي مجلة أخرى.

١٠- الناشر غير مسؤول عن محتوى ما ينشر من مادة الموضوعات التي تنشر في المجلة

١١- تكتب المراجع ضمن النص على الشكل التالي: [١] ثم رقم الصفحة ويفضل استخدام التهميش الإلكتروني المعمول به في نظام وورد WORD حيث يشير الرقم إلى رقم المرجع الوارد في قائمة المراجع.

تكتب جميع المراجع باللغة الانكليزية (الأحرف الرومانية) وفق التالي:

آ . إذا كان المرجع أجنبياً:

الكنية بالأحرف الكبيرة ـ الحرف الأول من الاسم تتبعه فاصلة ـ سنة النشر ـ وتتبعها معترضة (–) عنوان الكتاب ويوضع تحته خط وتتبعه نقطة ـ دار النشر وتتبعها فاصلة ـ الطبعة (ثانية . ثالثة) . بلد النشر وتتبعها فاصلة . عدد صفحات الكتاب وتتبعها نقطة.

وفيما يلي مثال على ذلك:

–MAVRODEANUS, R1986– **Flame Spectroscopy**. Willy, New York, 373p.

ب . إذا كان المرجع بحثاً منشوراً في مجلة باللغة الأجنبية:

ـــ بعد الكنية والاسم وسنة النشر يضاف عنوان البحث وتتبعه فاصلة، اسم المجلد ويوضع تحته خط وتتبعه فاصلة ـــ المجلد والعدد (كتابة مختزلة) وبعدها فاصلة ـــ أرقام الصفحات الخاصة بالبحث ضمن المجلة.

مثال على ذلك:

BUSSE,E 1980 Organic Brain Diseases **Clinical Psychiatry News** , Vol. 4. 20 – 60

ج. إذا كان المرجع أو البحث منشوراً باللغة العربية فيجب تحويله إلى اللغة الإنكليزية و التقيد بالبنود (أ و ب) ويكتب في نهاية المراجع العربية: (المراجع In Arabic)

رسوم النشر في مجلة جامعة البعث

١. دفع رسم نشر (٥٠٠٠٠) ل.س أربعون ألف ليرة سورية عن كل بحث لكل باحث يريد نشره في مجلة جامعة البعث.
٢. دفع رسم نشر (٢٠٠٠٠٠) ل.س مئة الف ليرة سورية عن كل بحث للباحثين من الجامعة الخاصة والافتراضية .
٣. دفع رسم نشر (٢٠٠) مئتا دولار أمريكي فقط للباحثين من خارج القطر العربي السوري .
٤. دفع مبلغ (١٥٠٠٠) ل.س ستة آلاف ليرة سورية رسم موافقة على النشر من كافة الباحثين.

المحتوى

اسم البحث	اسم الباحث	الصفحة
تحليل حركة المرور ضمن المدن الذكية	خالد قندقجي د.ماهر عباس د. أسماء شعار	١١-٤٨
تحسين بُنية الشبكات المُعرّفة برمجيّاً وحيدة المُتحكّم باستخدام بُنية شبكات البيانات المُسمّاة	علي مصطفى د.يمان غازي	٤٩-٧٨
تكامُل وتحسين الشبكات المُعرّفة برمجيّاً المُوزّعة باستخدام بُنية شبكات البيانات المُسمّاة	علي مصطفى د.يمان غازي	٧٩-١٠٢
تطوير PopScore كمقياس يعتمد على تحليل التشابه لتقييم نجاح المنشورات التسويقية على منصة الفيسبوك	راما أكرم ياسين د. ناصر أبو صالح د.وسيم رمضان	١٠٣-١٤٤

تحليل حركة المرور ضمن المدن الذكية

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

الملخص

باتت خوارزميات تعلم الآلة جزءً 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 المُعتمِدة 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 قندقجي 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.

د. أسماء شعار مدرس في كلية الهندسة المعلوماتية جامعة البعث قسم هندسة البرمجيات ونظم المعلومات

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 جزءً 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 (Intelligent Transportation Systems) ITS وغيرها العديد من التطبيقات.

تعاني معظم الدول وبشكل متزايد من مشكلة الازدحام المروري وبناءً على تقرير منظمة الصحة العالمية (World Health Organization) WHO ١,٣ مليون شخص يموتون كل سنة جراء حوادث مرورية وحوالي ٢٠-٥٠ مليون آخرين يعانون من أصابات 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 (Internet of Things) IoT 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 الصنعي 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:

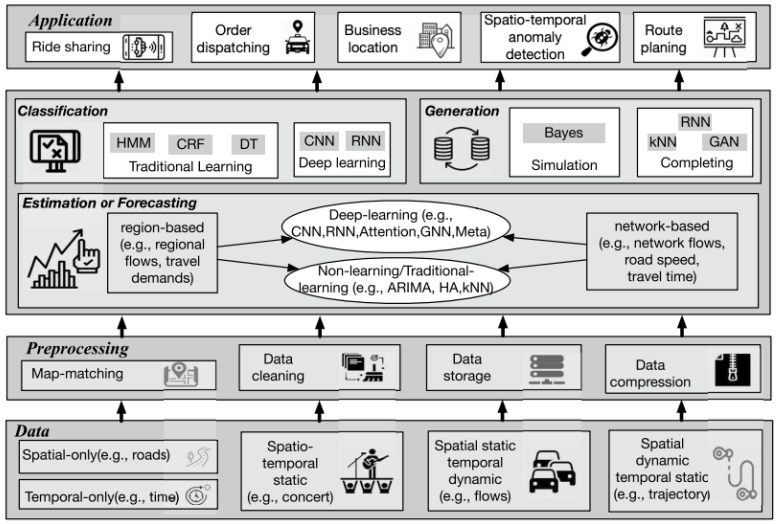
يهدف البحث للتنبؤ بحركة المرور ضمن المدن الذكية ودراسة تأثير بعض الخواص عليها كسرعة الحركة وقيم الانبعاثات الغازية كـ (CO_2, SO_2) التي تمثل حالة التلوث ضمن الجو بشكل مستقل عن القراءات السابقة أي بالاعتماد على القيم اللحظية التي تم قراءتها من الحساسات، بالإضافة لما سبق يهدف البحث أيضاً لدراسة تأثير قراءات 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 لزمن 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]

5-1 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.

5-2 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 كالبنى الموزعة ومنها (Apache spark and Hadoop) التي توفر إمكانيات عديدة كمعالجة تدفقات البيانات، التحليل السريع وسهولة الاستخدام [28].

5-3 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).

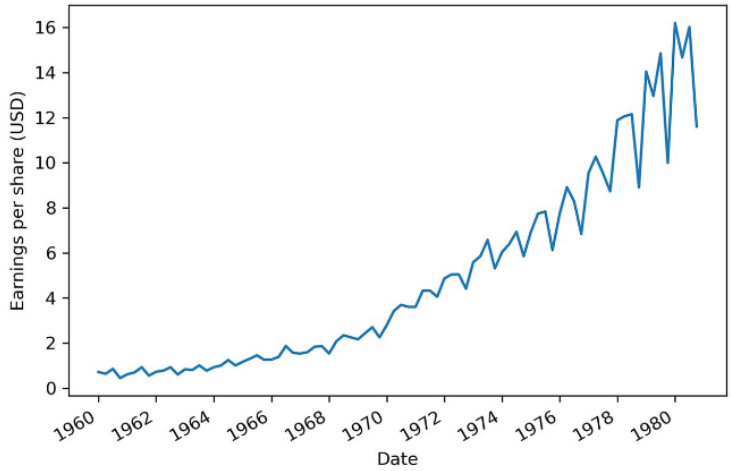
٢- التنبؤ (Prediction): وفيها نستخدم هذا النوع من الخوارزميات للتنبؤ مثلاً بحركة المرور ضمن طريق ما، ونستخدم في هذه الحالة البيانات التي يتم تجميعها على النقاط المثبتة ضمن الطرق. ومثال آخر التنبؤ بزمن تنقل السكان لأجل تحسين عملية النقل باستخدام سيارات الأجرة بحيث يتم نقل السكان بوقت مقبول ودون تأخير.

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

تُساهم عملية تحليل البيانات بتقديم العديد من التطبيقات المفيدة لتساعد في تحسين حركة المرور ضمن المدينة كاكتشاف 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 خطاً مستقيماً يوضح اتجاه السلسلة الزمنية.

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

٤- الخطأ (Residual): يمثل ما تبقى من السلسلة وتعبر عن الأخطاء الحاصلة وتسمى بـ (White Noise). وهذه المُركبة 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 و١٠ للاختبار. وأظهرت النتائج أن استخدام الخوارزميات الجينية قد حسن نتائج التدريب مقارنة باستخدام خوارزمية SVR بشكل عادي وحققت 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:

8-1 مجموعة البيانات المستخدمة في الدراسة:

تم استخدام مجموعة البيانات من مشروع citypulse EU FP7 [17] لمدنية 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 (CO_2, NO_2, ...) 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 (ttl.) وهي عبارة ملفات البيانات السابقة موسمه دلالياً 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 لمعرف يخص الطريق المدروس، كما ويوضح الجدول [2] عينة من بيانات التلوث المستخدمة والتي يتم جمع قراءتها أيضاً باستخدام الحساسات:

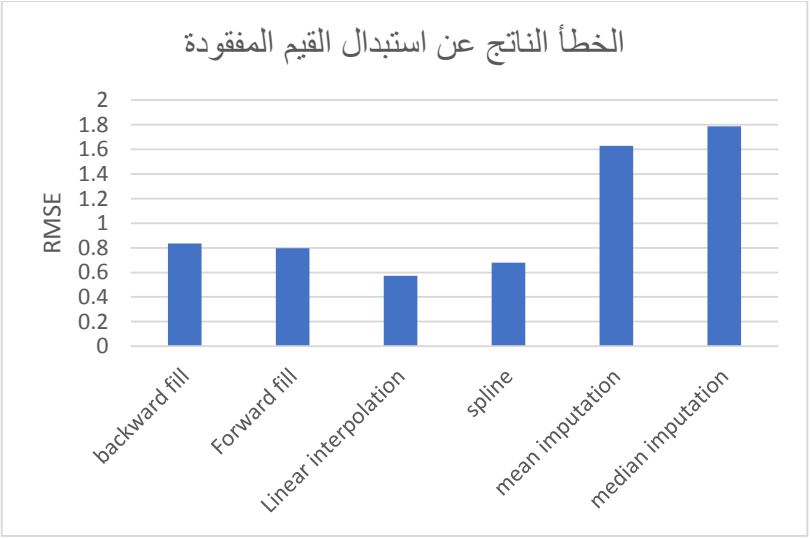
الجدول (٢) عينة من مجموعة البيانات التلوث المستخدمة [17]

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.

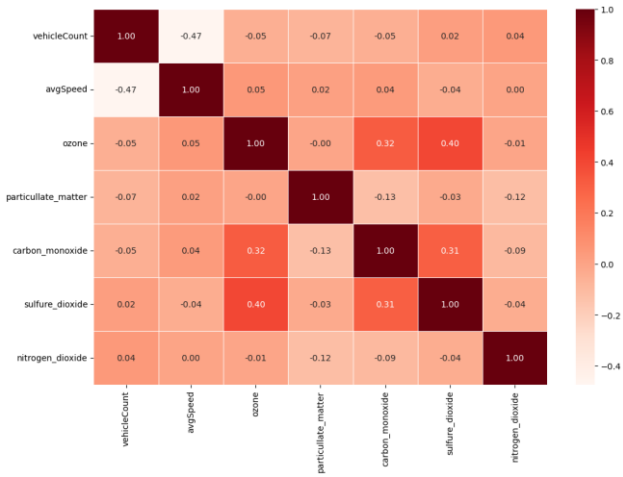
8-2 تجهيز البيانات Data Preparation:

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 كالمتوسط، أو الوسيط لكن مثل هذه الطرق لا تناسب طبيعة السلاسل الزمنية ويسبب تشوهاً للسلسة خاصة إذا كانت تتمتع بخاصية الموسمية (الفصلية) [25]. سنقوم بمقارنة الطريقتين السابقتين مع طرق أخرى تستخدم لاستبدال القيم المفقودة ضمن السلاسل الزمنية كالاستبدال بأخر قيمة سابقة (Forward Fill)، أو الاستبدال 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، وبالاستفادة من البيانات الفعلية والبيانات الناتجة عن عملية الاستبدال قمنا بحساب الخطأ الناتج عن هذه العمليات كما يوضح الشكل [3]:



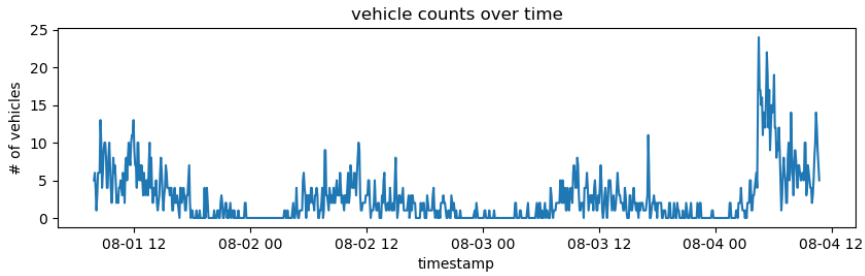
شكل (3) الخطأ الناجم عن استبدال القيم المفقودة للطرق السابقة.

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.



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

8-3 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 t−1	SO_2 t−1	CO_2 t−1	PM t−1	O_3 t−1	avgSpeed t−1	NO_2 t	SO_2 t	CO_2 t	PM t	O_3 t	avgSpeed t	**vehicleCount (output)**
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**

8-4 الخوارزميات وعوامل التقييم المستخدمة في الدراسة:

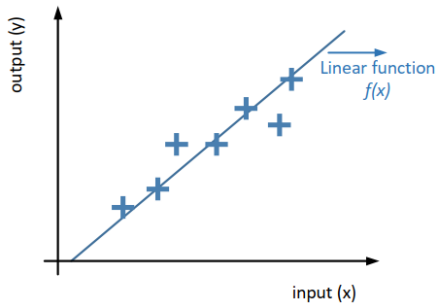
تستخدم خوارزميات الانحدار للتنبؤ بالقيم المستمرة (continuous values) وذلك من خلال إيجاد علاقة بين خواص الدخل والخرج. تدعم مكتبة تعلم الآلة sklearn (Science toolkit) [19] العديد من الخوارزميات التي يمكننا استخدامها نذكر منها:

8-4-1 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.

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

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] مثالاً 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.

$$y = m_1x_1 + m_2x_2 + \cdots + p \qquad (2)$$

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]:

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

حيث تعبر (n) عن عدد سجلات الدخل، (p) عدد الخواص المدروسة والمستخدمة 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:

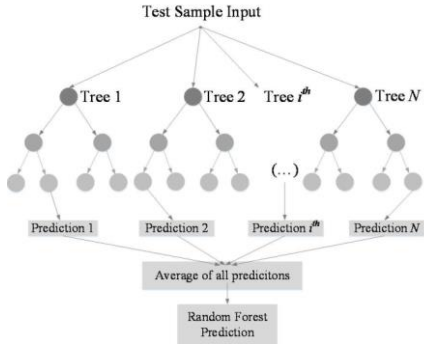
$$\widehat{\beta_{EN}} = \arg\min\{\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \alpha(\beta\sum_{j=1}^{p}|\beta_j| + (1-\beta)\sum_{j=1}^{p}\beta_j^2\} \qquad (4)$$

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 خاصية 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 والمرور 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]:

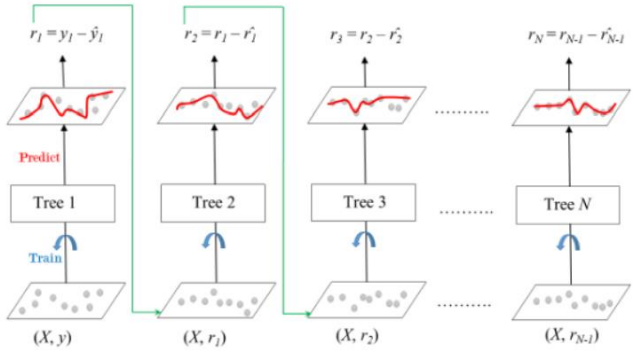


الشكل (7) التنبؤ بالقيم المستمرة باستخدام الغابات العشوائية [22]

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.

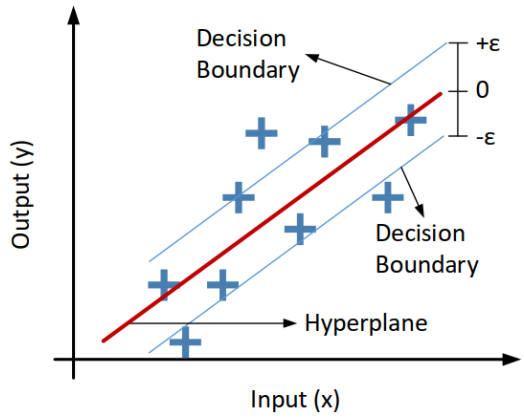


الشكل (8) التنبؤ بالقيم المستمرة باستخدام الانحدار التدريجي [23]

يتم ضبط هذه الخوارزمية بشكل مشابه لخوارزمية الغابات العشوائية، وتحوي 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 الجزر التربيعي لمتوسط الفروق بين القيم الحقيقية (O_i) والقيم الناتجة عن التنبؤ (P_i) ويعطى بالمعادلة [5]:

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

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.

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

أخيراً يمكن استخدام معامل R2 ويعطى بالمعادلة التالية:

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

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

الخوارزمية	المعاملات
Ridge Regression	`alpha=0.005`
ElasticNet Regression	`alpha=0.005, l1_ratio=0`

Decision Tree	`random_state=42, max_depth=3`
	`min_samples_split=2`
	`ccp_alpha = 0, splitter = 'best'`

Random Forest	`random_state=42, max_depth=10`
	`n_estimators=20`
	`min_samples_split=5`

Gradient Boosting	`random_state=42, max_depth=2`
	`n_estimators=20, min_samples_split=2`
	`learning_rate = 0.3`

SVR	`kernel='rbf', C=1`

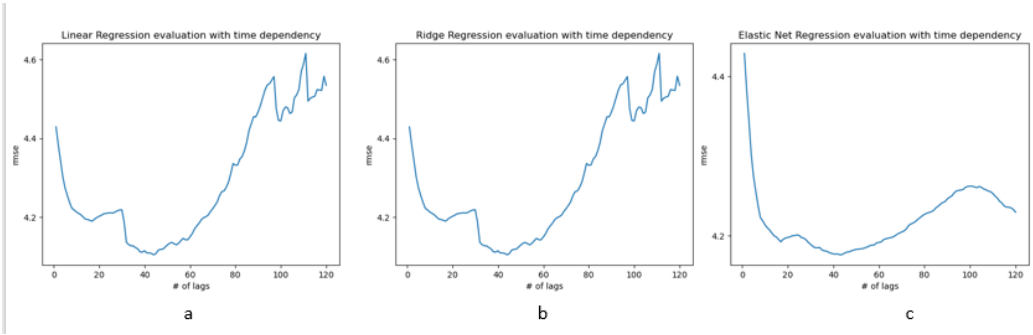
5-8 تأثير حجم النافذة (Window Size) على عملية التنبؤ:

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.

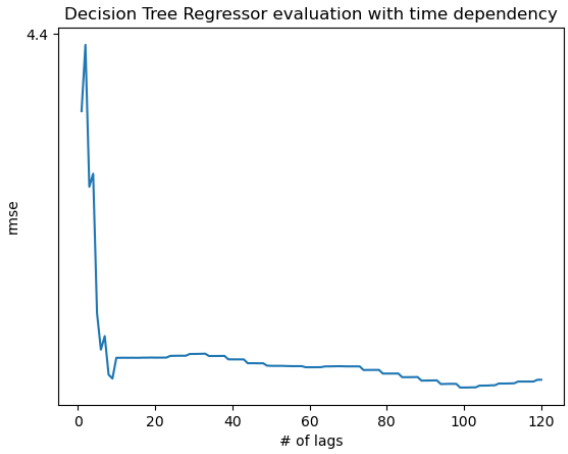
الخوارزمية	أفضل قيمة للنافذة	قيمة الخطأ (RMSE)
Linear Regression (LR)	44	4.104438
Ridge Regression	٤٤	٤,١٠٤٤٣٨
ElasticNet Regression (EN)	43	4.175865
Decision Tree	99	٤,٢٧٧٦٥٩
Gradient Boosting (GB)	104	3.827373
Random Forest (RF)	115	3.630251
SVR	104	٣,٨٢٧٣٧٣

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).

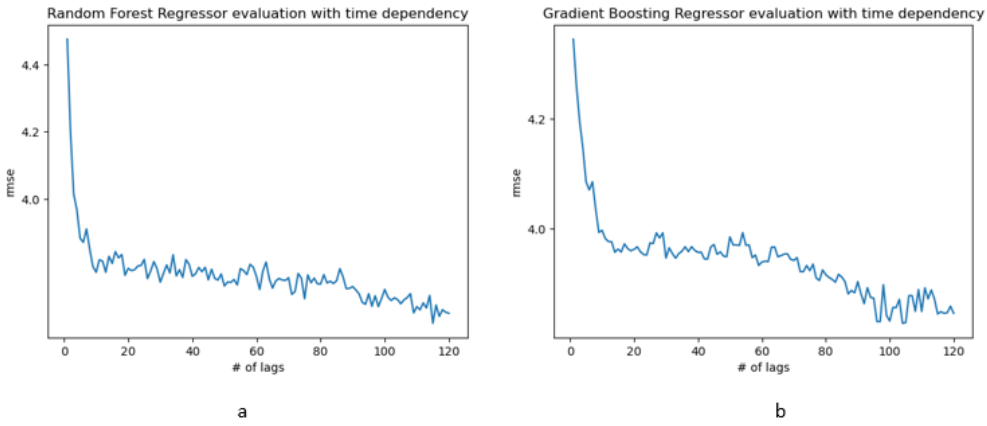


الشكل [10] تأثير حجم النافذة على خوارزميات التنبؤ: (a): LR, (b): ridge, (c):(EN)

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.

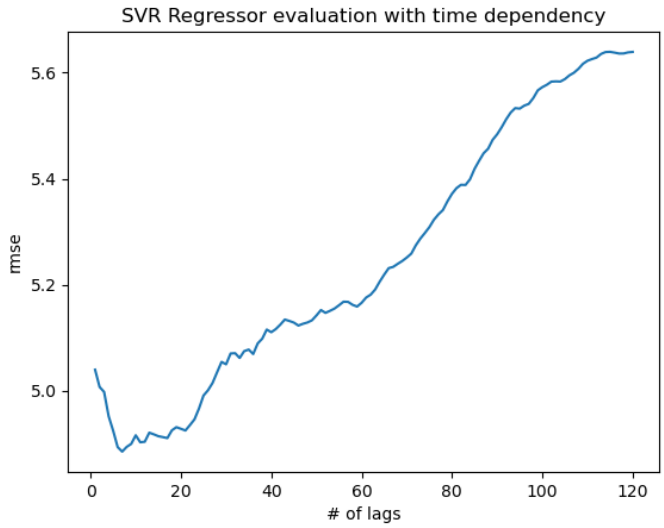


الشكل [11] تأثير حجم النافذة على خوارزمية أشجار القرار



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.



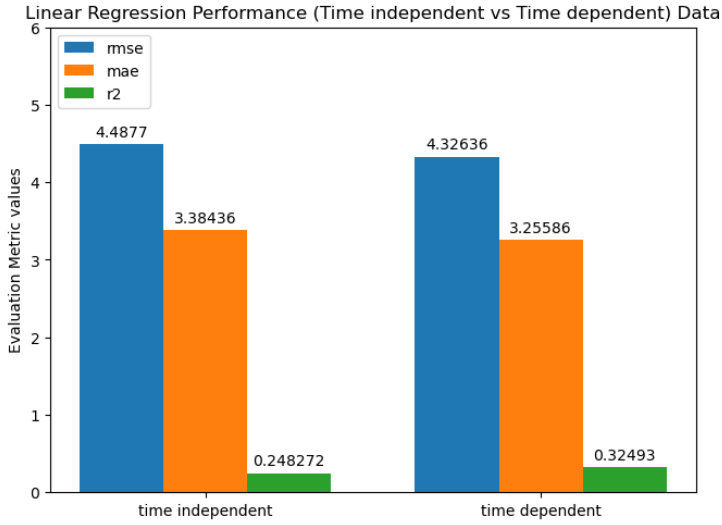
الشكل [13] تأثير حجم النافذة على خوارزمية SVR.

6-8 تأثير استخدام القراءات السابقة في عملية التنبؤ:

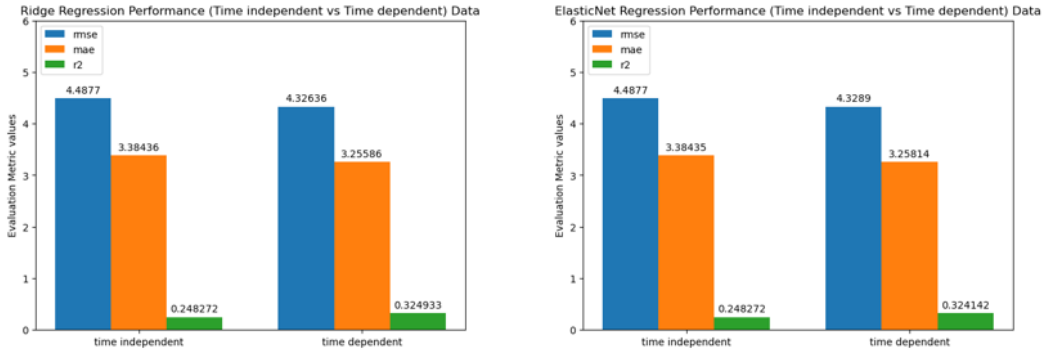
تحتاج معظم الخوارزميات لحجم نافذة 44 على الأقل حتى نبدأ بالحصول على أقل معدل خطأ ممكن ويتحسن أداء عملية التنبؤ، لذا ولمقارنة تأثير استخدام القراءات السابقة سنقوم بحساب متوسط الخطأ الناتج عن استخدام القراءات السابقة للحساسات من حجم النافذة 44 وحتى 120 ومقارنة نتائج الخطأ مع استخدام القراءات الحالية للحساس.

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 r2 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.

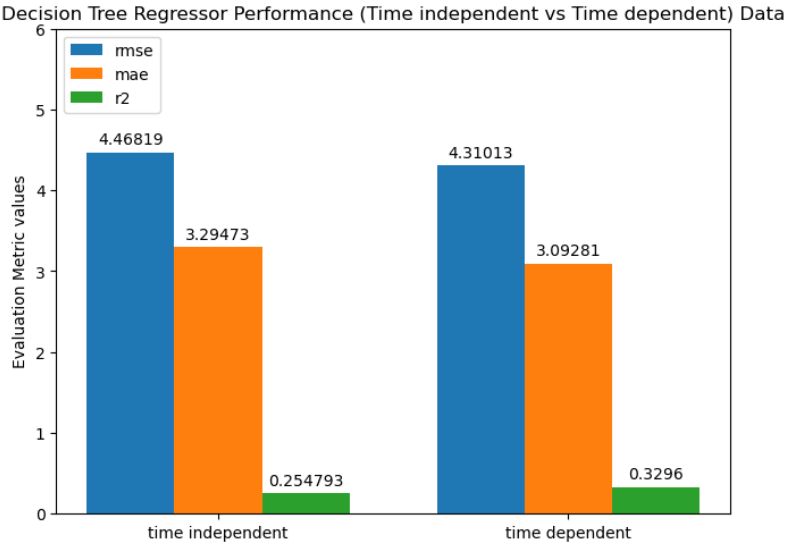


الشكل [14] تقييم أداء خوارزمية (LR).



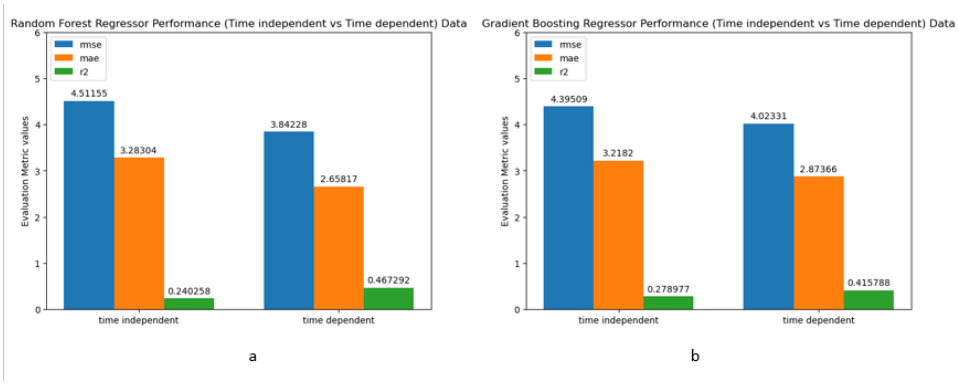
الشكل [15] تقييم أداء خوارزميتي: EN :(b), ridge :(a)

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



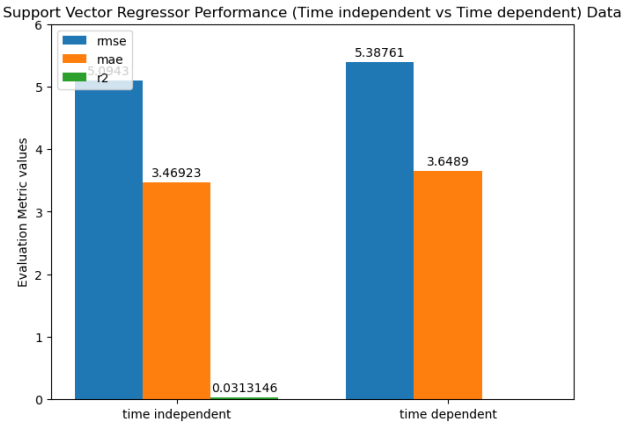
الشكل [16] تقييم أداء خوارزمية (DT).

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: GB :(b), RF :(a)

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



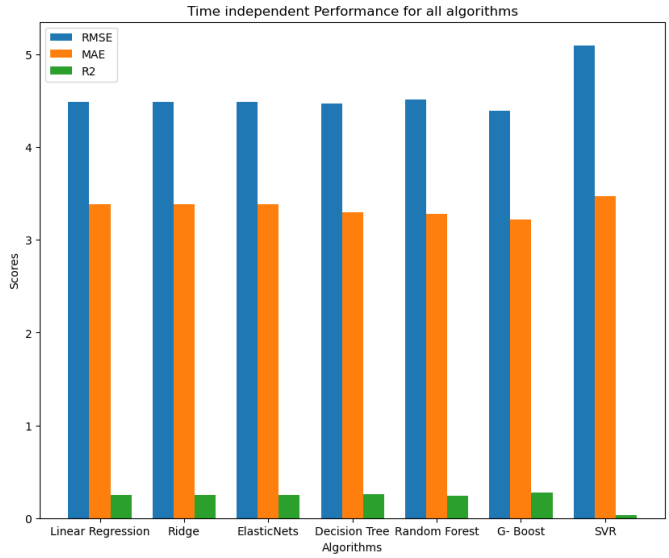
الشكل [18] تقييم أداء خوارزمية (DT).

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

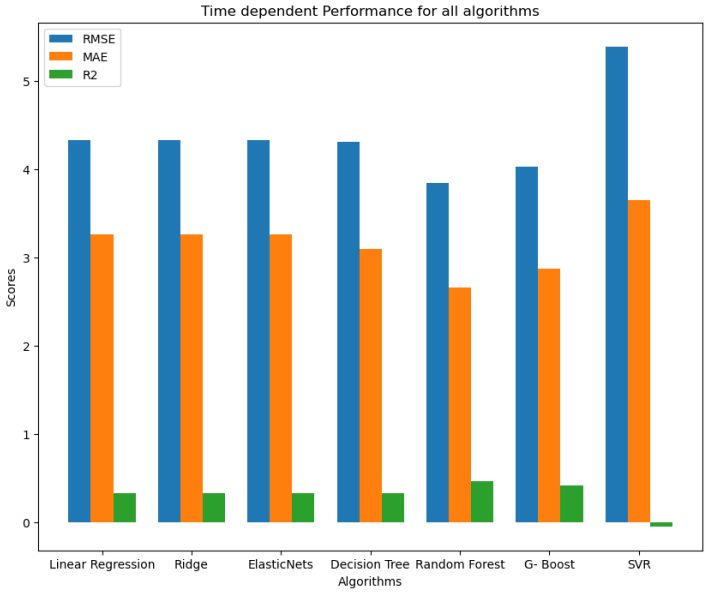
الجدول (7) تقييم أداء الخوارزميات المدروسة.

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 التجربتين المدروستين.

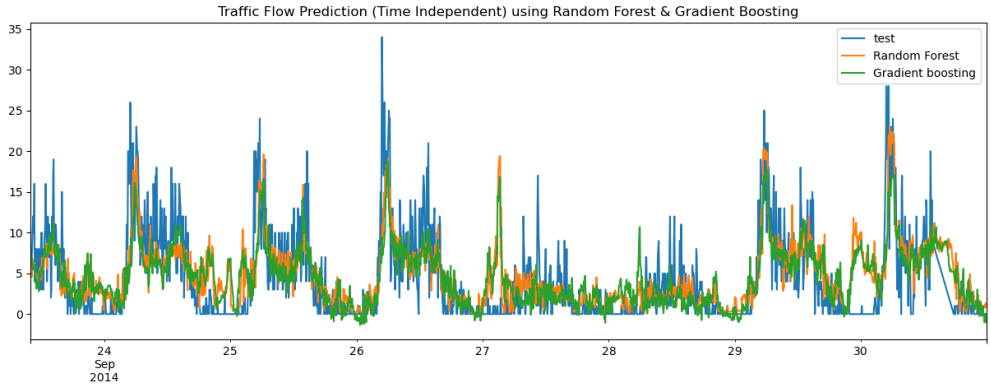


الشكل [19] تقييم أداء الخوارزميات المدروسة عند استخدام القراءات الحالية للحساس.



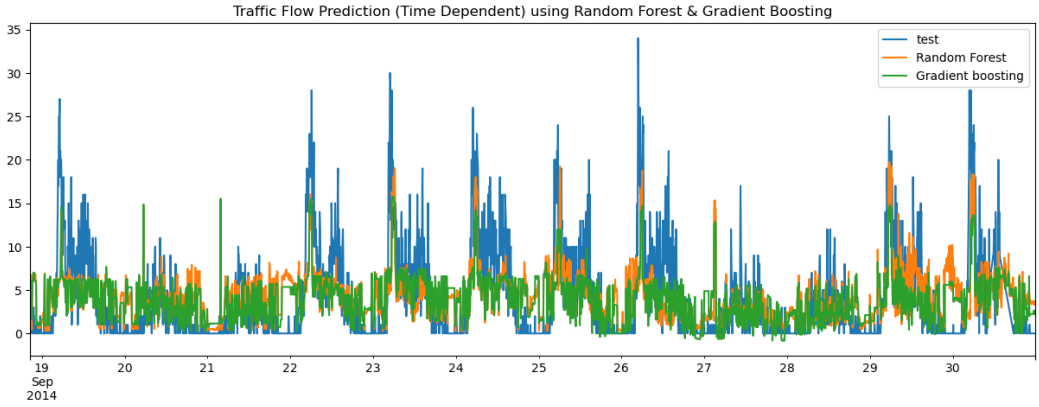
الشكل [20] تقييم أداء الخوارزميات المدروسة عند استخدام القراءات الحالية والسابقة للحساس.

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

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.

١١- المراجع:

[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).

تحسين بُنية الشبكات المُعرّفة برمجيّاً وحيدة المُتحكّم باستخدام بُنية شبكات البيانات المُسمّاة

د.م. يمان غازي[١] م. علي مصطفى[٢]

المُلخّص

ظهر حديثاً تكامُل للشبكات المُعرّفة برمجيّاً (SDN) وشبكات البيانات المُسمّاة (NDN) ضمن بُنية جديدة عملت على تحسين توزّع المحتوى في الجيل الحديث من بُنى الشبكات. حيث تُوفّر الشبكات SDN مستوى تحكّم مركزيًا يُحسّن قرارات التوجيه؛ كما إنَّ الميّزات ضِمن مستوى بيانات شبكات NDN، مثل التسليم المُعتمد على الاسم، والتخزين المؤقت داخل الشبكة، والتوجيه المُتكيّف، تُسهّل من آلية نشر البيانات والوصول إليها.

تهدُف هذه الدراسة إلى تحقيق البنية الجديدة SD-NDN ذات المُتحكّم الوحيد في بنية المحاكي ndnSIM وسنقوم بذلك عن طريق كتابة الترميز المصدري المُحاكي لعملها وتضمينها في المُحاكي وأيضاً تحقيق البنية IP-Based المُشابهة لها ضمن المحاكي Mininet من أجل المُقارنة بين البُنيتين في البيئة NDN-Based والبيئة IP-Based حيث سيتم اسقاط ذلك من خلال إضافة تحديد طوبولوجيا مُوحّدة لكل سيناريو وتحديد وسطاء تجربة مُناسبة، ومن ثم سيتم التحقيق خلال تنفيذ سيناريو عملي على كلا المُحاكيين، وفي الجُزء الأخير من الدِراسة سيتم تقييم أداء البُنى المُحقّقة ومُقارنة كل شبكة IP-Based مع مثيلتها ضمن البيئة NDN-Based. أثبتت النتائج تفوّق البُنى ضمن بيئة NDN من ناحية عدِة وسطاء أداء مثل الإنتاجية وعدد الرُزم المفقودة والتأخير مُقارنةً مع البيئة التقليدية IP-Based، بينما كان استخدام عرض الحزمة أكبر في البُنى NDN-Based.

[١] استاذ مُساعد، قسم هندسة الشبكات والنظم الحاسوبية، كلية الهندسة المعلوماتية، جامعة حمص
[٢] طالب دراسات عليا (ماجستير)، قسم هندسة الشبكات والنظم الحاسوبية، كلية الهندسة المعلوماتية، جامعة حمص

الكلمات المفتاحية: شبكات البيانات المُسمّاة، الشبكات المُعرّفة برمجيّاً وحيدة المُتحكّم، الشبكات المركزيّة، إنترنت المستقبل، شبكات SDN NDN (SD-NDN).

Improving Single -Controller Software-Defined Networks Architecture Using Named Data Networks Architecture

Eng. Ali Mustafa[٤] Dr. Eng. Yaman Ghazi[٣]

ABSTRACT

Recently, software-defined networking (SDN) and named data networking (NDN) have merged into a new architecture that improves content distribution in the next generation of network architectures. SDN provides a centralized control plane that improves routing decisions, while features within the NDN data plane, such as name-based delivery, in-network caching and adaptive routing, facilitate data dissemination and access. This study aims to implement the new single-controller SD-NDN architecture in the ndnSIM simulator. We will do this by writing the simulated source code and including it in the simulator. We will also implement a similar IP-based architecture in the Mininet simulator to compare the two architectures in an NDN-based and IP-based environment. This will be achieved by adding a unified topology to each scenario and specifying appropriate experimental parameters. The implementation will then be executed through a practical scenario on both simulators. The final part of the study will evaluate the performance of the

[3] Assistant Professor, Department of Network Engineering and Computer Systems, Informatics Engineering Faculty, Homs University.

[4] Postgraduate Student (M.A), Department of Network Engineering and Computer Systems, Informatics Engineering Faculty, Homs University.

implemented architectures and compare each IP-based network with its counterpart in an NDN-based environment. The results proved that the NDN environment outperformed the IP-based architectures in terms of several performance parameters, such as throughput, number of packets lost, and delay, compared to the traditional IP-based environment, Bandwidth utilization was higher in the NDN-based architectures.

Keywords: Named Data Networking, Software-Defined Networks, Information Centric Networks, Future Internet, SDN-NDN Single-Controller Networks (SD-NDN).

١. مُقدمة:

تعتمد بُنية عُنق الزجاجة للإنترنت (Hourglass Internet Architecture) اليوم على طبقة الشبكة العالمية الحالية Internet Protocol (IP). لقد تم تصميم شبكات الـ IP في الأصل للاتصال بين نقطتي نهاية. الآن ومع النمو الهائل والمستمر للتجارة الإلكترونية والشبكات الاجتماعية والوسائط الرقمية وتطبيقات الهواتف الذكية، يهيمن توزيع المحتوى واسترجاعه بشكل متزايد على الاتصال عبر الإنترنت. هذا الاستخدام المهيمن للإنترنت مثل شبكة التوزيع تسبب مشاكل التوزيع. يُعتبر حل مشاكل التوزيع في الشبكة المُوزّعة عبر بروتوكول الاتصال من نقطة إلى نقطة(IP-Based) عُرضَة للخطأ ومُعقد. تقوم المُوجّهات في مستوى التوجيه بتبادل تحديثات التوجيه واختيار أفضل المسارات لإنشاء جدول إعادة التوجيهforwarding Information Base (FIB) ، تقوم أجهزة التوجيه بإعادة توجيه الرُزم باستخدام هذا الجدول FIB. [1]

إن البروتوكول IP لديه نهج توجيه ذكي ونهج إعادة توجيه، ويتم استخدام شبكات البيانات المُسمّاة Named Data Networks (NDN) لتعزيز أسلوب إعادة التوجيه. توجد بُنى إنترنت مستقبلية حالية كثيرة وأشهرها الشبكات المُرتكِزة على المُحتوى Content-Centric Networks (CCN)، وإنّ المفهوم الرئيسي وراء CCN هو أنهُ يتم انشاء الطلب من قِبل طالب البيانات أو المُستهلك (Consumer) حتى يتم بعدها توجيه ذلك

الطلب في اتجاه ذلك الموقع في الشبكة الذي تتُوفّر فيه هذه البيانات. يتم اجتياز بعض العُقد التي على المسار المؤدي إلى مُنتج البيانات المطلوبة (Producer). تحتوي هذه العُقد على ذاكرات تخزين مؤقت (Cash) يتم فحصها أثناء النقل، حيث من الممكن وجود نسخ من المعلومات المطلوبة من قبل المُستهلك عند تلك العُقد الوسيطة. [1]

قُدّمت الكثير الدِراسات عن طريقة إدارة شبكات SDN لعمليات التوجيه والتدفّق ضِمن مُبدّلات SDN التقليدية، ولكن لن يُفيدنا هذا الأمر بتاتاً في بُنية الشبكات SDN–NDN Networks (SD–NDN) المُتكاملة نظرًا لوجود قواعد توجيه قائمة على الاسم (Name-Based Forwarding). يمكن لعُقد NDN التعرف على طلبات مختلفة لنفس المحتوى أو توفير البيانات مباشرةً من المخزن المؤقت CS الخاص بها دون اللجوء إلى المُتحكّم وانتظار قراراته. لذلك، يمكن للوسطاء المُتعلّقة بالمحتوى، مثل تواتر الطلب وشعبيته، أن تؤثر على حالة إشغال قاعدة بيانات FIB وإدارتها، ونسبة التفاعلات مع المُتحكّم. حتى الآن، قامت أغلب الأبحاث في مجال دمج وتكامل بُنيتي SDN وNDN بشكل أساسي على تصميم البُنية المعماري وتقييم أداء توصيل المحتوى دون النظر في تأثير إعادة التوجيه المستند إلى الاسم على أداء FIB ودون النظر والتفصيل بإمكانيّة دمجها مع شبكات SDN وتحديد التحسينات. [2]

٢. الهدف من البحث:

إنّ المُساهمة الأساسيَة في هذا البحث هي تطبيق بنية شبكات NDN على بنية SDN ذات المُتحكّم الوحيد ضمن المُحاكي ndnSIM وذلك بكتابة الترميز المصدري الذي يحاكي عمل هذه البُنى وتضمينها في المحاكي ثم استخدام المُحاكي Mininet لمُحاكاة نفس هذه البُنى. وتحديد أهم مُعامِلات الأداء، وتحديد التحسينات التي سنحصل عليها من ناحية البُنية عند تحقيق هذا الاندماج.

٣. الشبكات المُعرّفة برمجيّاً (Software-Defined Networks)

تحوي الشبكات الحاسوبيّة عدد كبير من أجهزة الشبكة، مثل أجهزة التوجيه والمُبدّلات وأنواع عديدة من الأجهزة التي تتعامل مع المرور لأغراض أخرى غير توجيه الرُزم، مع العديد من البروتوكولات المعقدة التي يتم تنفيذها عليها. إنّ مُشغّلوا الشبكة ومُديروها مسؤولون عن إعداد السياسات بقصد إعداد مجموعة واسعة من تطبيقات وأحداث الشبكة، وعليهم تحويل هذه الإعدادات عالية المستوى يدوياً إلى أوامر مُنخفضة المستوى للتكيّف مع ظروف الشبكة المُتغيّرة، ولذلك تك اقتراح فِكرة "الشبكات القابلة للبرمجة" لتجاوز السلبيات الموجودة في الشبكات التقليدية، حيث تم تطوير الشبكات المُعرَّفة برمجيّاً (SDN) للحُصول على تحكُّم مُستقل ولتسهيل التطوير والابتكار وتمكين التحكُّم البرمجيّ البسيط لمسار (أو مسارات) بيانات الشبكة. إن فصل المُوجّهات عن منطق التحكُّم يسمح بنشر أسهل للبروتوكولات والتطبيقات الجديدة، وتصوّر وإدارة الشبكة بشكل واضح، فبدلاً من إعداد وتطبيق السياسات وتنفيذ البروتوكولات على الأجهزة المُتناثرة جغرافيّاً، يتم اختصار الشبكة إلى أجهزة توجيه بسيطة ووحدة تحكُّم (أو وحدات تحكُّم) بالشبكة هي من تقوم بصُنع القرار ونشره إلى جميع الأجهزة الشبكيّة المسؤولة عنها. [3]

٤. شبكات البيانات المُسمّاة (Named Data Networks)

إنّ شبكة البيانات المُسمّاة (NDN) هي بُنية شبكات مركزيّة للمعلومات (ICN) تقوم على تحويل نموذج الإنترنت المعروف حاليّاً من نموذج مُتمركز على المُضيف (host-centric) إلى نموذج مُتمركز على المعلومات (information-centric)، مع استخدام أسماء مُحتوى التطبيق مباشرةً في طبقة الشبكة للبحث عن البيانات واسترجاعِها.

تُعدُّ هذه الشبكات بُنية إنترنت جديدة تمامًا مُستوحاة من سنوات من البحث التجريبي في استخدام الشبكة. ترتبط NDN بالشبكات التي تركز على المحتوى (Content Centric Networking). الميزة الفريدة لـ NDN هي مستوى التوجيه المُتكيّف (adaptive forwarding plane). حيث تحمل الرُزم في شبكات NDN اسم البيانات بدلاً من عنوان المصدر والوجهة. يتم في شبكات NDN الاتصال عن طريق تبادل رُزم البيانات Data

والطلب Interest، حيث يُرسل مستهلكوا البيانات (Consumers) رُزم Interest على شكل أسماء.

تقوم أجهزة التوجيه بإعادة توجيه رُزمة Interest بِناءً على اسم البيانات وتُحافظ أيضًا على معلومات حالة الاهتمامات المُعلّقة التي تُمكّن أجهزة توجيه NDN من اكتشاف الحلقات وقياس أداء المسار المختلف واكتشاف حالات الفشل بسرعة وإعادة المحاولة عبر مسار بديل. يردُّ المُنتِج (Producer) برُزمة بيانات تأخذ مسار طلب الرسالة Interest العكسيّ.

ترِثُ شبكات NDN شكل بُنية الساعة الرمليّة لمعمارية شبكات IP ولكنه يستبدل طريقة تبادل البيانات عبر العنوانين IP بنموذج استرجاع بيانات (data retrieval model) عند الخصر النحيف للساعة الرمليّة. [1]

٥. البُنية الجديدة SD-NDN:

إنّ البُنيتين في شبكات NDN وشبكات SDN تهدُفان إلى التغلّب على قيود بنية الإنترنت الحالية مثل ضعف الأداء، والتأخير، وفقدان الرُزم، وضعف الأمان، وضعف قابلية التوسع. يُعدّ NDN نهجًا يُركّز على المحتوى، حيث يستخدم أسماءً يسهل على البشر قراءتها لطلب البيانات وتسليمها، بغض النظر عن مكان تخزينها أو استضافتها. بالإضافة إلى ذلك، يُعزّز الأمان بفضل الأمان المُدمج الذي يتحقق من سلامة البيانات وأصالتها ومصدرها من قِبل أي عقدة في الشبكة، دون الحاجة إلى آليات خارجية مثل الشهادات أو التشفير. على النقيض، يعتمد نهج SDN على الفصل بين مستوى التحكم ومستوى البيانات في أجهزة الشبكة، حيث تتولى وحدة تحكم مركزية مسؤولية إدارة وظائف الشبكة والتحكم بها، بينما تقتصر مهام الأجهزة على إعادة توجيه الرُزم وفقًا للتعليمات الصادرة عن وحدة التحكم. يوفّر هذا النهج العديد من المزايا مقارنةً بالأسلوب الموزّع التقليدي، إذ يُحسّن أداء الشبكة وكفاءتها واستقرارها، مستفيدًا من رؤية شاملة تُمكّن وحدة التحكم من مراقبة حالة جميع الأجهزة داخل الشبكة وتوظيف هذه المعلومات لتعزيز أدائها. بالإضافة إلى ذلك، يوفّر هذا

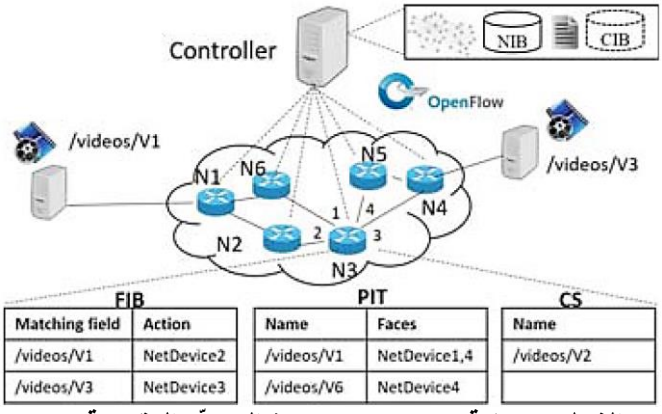
النموذج مرونة عالية وقابلية للتوسع، بفضل التكوين الديناميكي الذي يسمح لوحدة التحكم بتكييف أداء الشبكة في الوقت الفعلي وفقًا للظروف والمتطلبات المتغيرة. [4]

يتيح دمج شبكات SDN و NDN تحسين آلية التوجيه في NDN عبر SDN، وذلك من خلال إدارة قاعدة معلومات إعادة التوجيه (FIB) الخاصة بشبكات NDN. حيث تقوم وحدة تحكم SDN بحساب مُدخلات FIB المُثلى استنادًا إلى رؤية شاملة لحالة الشبكة، ثم تُحدّث FIB ديناميكيًا لضمان التكيف مع التغييرات والملاحظات المستمرة. [4]

إلا أنّ هذا التكامل يفرض تحديات ومخاطر أمنية جديدة تستوجب المعالجة. فعلى سبيل المثال، قد يتمكن مهاجم من اختراق مُتحكّم SDN والتلاعب بإدخالات FIB لعُقد NDN، مما يؤدي إلى تعطيل عملية توصيل المحتوى، ويهدد توافر بيانات NDN وسلامتها وسريتها. علاوةً على ذلك، قد لا تتوافق إدخالات FIB لعُقد NDN مع الرؤية الشاملة لوحدة تحكم SDN، مما قد يتسبب في أخطاء توجيهية وانخفاض في الكفاءة، الأمر الذي يؤدي في النهاية إلى تدهور أداء بيانات NDN وموثوقيتها. [4]

إنَّ شبكات NDN مُعرّضة للهجمات التي تستغل ميزات بنية NDN، مثل إغراق الاهتمامات، أو تسميم المحتوى، أو اختطاف وتزوير التسمية، وهو ما يمكن أن يؤثر على توافر البيانات وسلامتها وصحتها عبر شبكة NDN. تجدُر الإشارة إلى أن SDN قد لا تتمكن من اكتشاف أو منع مثل هذه الهجمات لأنها فريدة من نوعها لبروتوكول NDN ودلالاته. بالتالي، يُشكّل الأمان عُنصرًا حاسمًا يجب أخذه بعين الاعتبار عند دمج شبكتي SDN و NDN. فمن الضروري تنسيق آليات الأمان في كلا النموذجين لضمان حماية بيانات NDN وضبط التحكم في SDN. إضافةً إلى ذلك، يتعيّن تطوير حلول أمنية مبتكرة للتصدي للتحديات والمتطلبات الفريدة التي تفرضها هذه البنية المدمجة. تُحدثُ الشبكات NDN و SDN تحولًا جذريًا في بنية الشبكات عبر إعادة تشكيل أسسها، حيث تنقل NDN الاتصال من نموذج قائم على المُضيف (IP-Based) إلى نموذج يعتمد على المحتوى (Content-Based)، في حين تُحوّل SDN آلية التحكم من التوزيع إلى المركزيّة. إن الاستفادة من نقاط القوة المتكاملة في كلا النموذجين يمكن أن تُسهم في تطوير بنية الإنترنت، بما يواكب احتياجات المستقبل وتحدياته. [4]

يُوضّح الشكل (١) البنية وحيدة المُتحكّم المُقترحة لشبكات SD-NDN وِفق المرجع [2]:



الشكل (١) بنية SD-NDN وحيدة المُتحكّم المُقترحة [2]

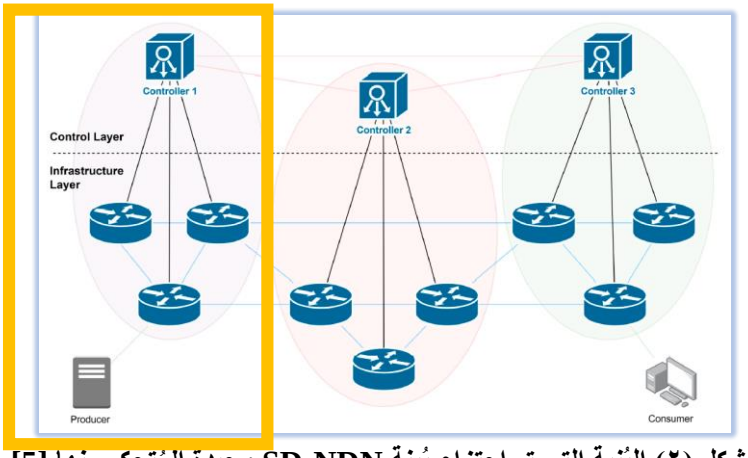
يجب الحِرص عند دمج SDN ضمن شبكات NDN على تجنّب التعديلات الجذرية، حيث إن التغييرات الكبيرة قد تُفقد NDN العديد من مزاياه الأساسية، مما قد يؤدي إلى جعل نموذج SD–NDN أقل كفاءة مقارنةً بنموذج NDN البحت أو النموذج المُعتمِد على المُضيف.

يوضّح الشكل (٢) مثالًا على بنية مقترحة لمواجهة هذه الإشكالية، حيث تُقسَّم الشبكة إلى طبقتين :طبقة التحكم وطبقة البنية التحتية .تتألف طبقة التحكم من مجموعة من عُقد التحكم، التي تُشرف على إدارة العديد من مُوجّهات NDN. ولتجنّب إرهاق وحدات التحكم، ينبغي أن يكون عدد مُوجّهات NDN المُخصصة لكل وحدة تحكم متناسبًا مع إمكاناتها وقدرتها التشغيلية.

أما طبقة البنية التحتية، فتتكوّن من مُوجّهات NDN إلى جانب المنتجين والمستهلكين. وعلى عكس شبكات NDN التقليدية، فإن عُقد التوجيه هنا تقتصر مسؤوليتها على إعادة توجيه الحزم وعمليات التبديل، بينما يُنقل عبء التوجيه إلى وحدات التحكم.

تُبرز هذه البنية الموضحة في الشكل (٢) نقطة أساسية، وهي أن العديد من بُنى SD– NDN قد تُؤدي بشكل غير مقصود إلى تقويض المزايا التي يُفترض أن يُحققها NDN،

مما قد يجعل نموذج SD–NDN أقل كفاءة من NDN الخالص أو من النهج القائم على المضيف. [5]



الشكل (٢) البُنية التي تم اجتزاء بُينة SD-NDN وحيدة المُتحكم منها [5]

سنقوم باعتماد جُزء من هذه البينة (الشكل ٢) ضمن البنية SD-NDN وحيدة المُتحكّم، وسنستخدم وسطاء خاصة بالمُحاكاة (سنُوضّحها لاحقاً ضمن **القسم 1.8**).

٦. أدوات البحث

قمنا في هذا البحث باستخدام المُحاكيين التاليين لتحقيق البنية العمليّة ضمن البيئتين IP-Based و NDN-Based:

1.6 مُحاكي الشبكات Mininet

يُعد Mininet برنامج محاكاة قائمًا على Linux يُستخدم للنمذجة السريعة للشبكات المعرفة برمجيّاً. يتيح تشغيل وإدارة مجموعة من الأجهزة المضيفة، والمبدّلات، والمُوجّهات، والوصلات، مستفيدًا من تقنيات المحاكاة الافتراضية لجعل نظام واحد يعمل وكأنه شبكة متكاملة. يتم إنشاء هذه العناصر، بما في ذلك المضيفات والمبدّلات ووحدات التحكم، عبر البرمجيات بدلًا من hardware، ومع ذلك، فإن سلوكها يُحاكي إلى حد كبير نظيراتها من مكونات الأجهزة الفعلية. [6]

يُوفّر المُحاكي Mininet طريقة بسيطة ومنخفضة التكلفة لاختبار الشبكات وتطوير تطبيقات OpenFlow، حيث يُتيح اختبار طوبولوجيا شبكية واسعة ومعقدة دون الحاجة إلى بنية تحتية مادية. كما يُمكّن المستخدم من التحكّم في الشبكة الافتراضية وإدارتها من Console واحد. يتضمّن Mininet واجهة سطر أوامر مُدرِكة للشبكة (CLI) التي تتيح إنشاء الشبكة واختبارها بمرونة وسرعة. [7]

2.6 مُحاكي الشبكات ndnSIM

تُعد بيئة المحاكاة ndnSIM أداةً بالغة الأهمية للباحثين المهتمين بدراسة شبكات NDN، إذ تُمكّنهم من تنفيذ تجارب المحاكاة، وتحليل نتائجها، وتقييم الأداء. تُوفر هذه البيئة إمكانيات واسعة، بما في ذلك إنشاء طوبولوجيا الشبكة، وتحديد معايير المحاكاة، وتنفيذ نموذج بروتوكول طبقة الاتصال، فضلًا عن محاكاة الاتصال بين عُقد NDN المختلفة، وتسجيل أحداث المحاكاة، وعرض النتائج بشكل مرئي.

يُعتبر ndnSIM برنامجًا مفتوح المصدر يعتمد على المُحاكي ns-3، مما يسمح بإجراء تجارب على بنية NDN في الشبكات السلكية واللاسلكية. يتميز ndnSIM بتكامله الكامل مع نماذج NDN في العالم الحقيقي، بما في ذلك ndn-cxx و NDN Forwarding Daemon (NFD).

كما يُوفّر ndnSIM مجموعة واسعة من التوابع المساعدة (Helpers)، التي تُتيح تتبّعًا تفصيليًا لسلوك كل مكوّن، بالإضافة إلى تدفق حركة مرور NDN. يستخدم المحاكي مكتبة ndn-cxx ومكتبة C++ مع توسعات تجريبية، إلى جانب NFD لضمان واقعية عمليات المحاكاة وإمكانية إعادة إنتاجها في بيئات حقيقية، دون الحاجة إلى تغييرات كبيرة في التعليمات البرمجية المصدر. يتم الاعتماد على صفوف C++ المختلفة لنمذجة سلوك كل كيان. [8]

تم تنفيذ تجارب وسيناريوهات المحاكاة باستخدام ndnSIM 2.9 (وهو الإصدار الحاليّ)، بعد تثبيته ضمن بيئة NS-3 على توزيعة Ubuntu Desktop 20.04 LTS من نظام التشغيل Linux.

٧. الدِراسات السابِقة:

هناك العديد من الدراسات التي تناولت موضوع التكامل بين شبكات NDN وشبكات SDN، وقد اتسم الكثير منها بالطابع النظري. تركّز غالبية الأبحاث على دراسة شبكات NDN وتحسينات التوجيه فيها، بما يشمل اختيار أفضل المسارات وتحديد وسطاء الأداء مثل التأخير، والتغيّرات، وعرض الحزمة وغيرها، ووجب التنويه هنا إلى قِلة المقالات الاختصاصيّة ذات الطابع العمليّ التفصيليّ حول البُنية SD-NDN لذلك حاولنا قدر الإمكان إنجاز وتحقيق سيناريوهات تتوافق مع المقالات المرجعية وما تحويه من توصيف للبنية ووسطاء.

غالبًا ما يتم مقارنة مختلف السيناريوهات استنادًا إلى وسطاء ذات صلة بعملية التوجيه، مما أدى إلى بعض التحديات في المراحل الأولى، خاصة عند تحديد سيناريوهات المحاكاة والمعايير المستخدمة في التقييم.

فمثلاً يقوم الباحثين في الورقة البحثية [2] بتطبيق بنية SDN-NDN باستخدام OpenFlowكواجهة Southbound Interface كما هو مُستخدم في المراجع [1] و [9]، حيث هنا تم استخدام المتحكّم لاتخاذ قرارات التوجيه بدلاً من الطريقة التقليدية ضمن NDNوتم الحصول على تحسينات في توجيه الطلبات. سنعتمد في هذه الورقة على دراسة هذه البنية واحتمال استخدامها لاحقاً ضمن أي توسعات مستقبلية مع بنية الشبكات SDN.

بينما اقترح الباحثون في الورقة البحثيّة [10] بنية وبروتوكول شبكة جديدين، وتم إعطاء كلاهما المسمى NDN-Fab، للسماح بالتفاعل بين بنى الشبكة المُعتمدة على العناوين (Ip-based)والمحتوى (Content-based) وتسمح NDN-Fab للشبكة الطرفية بالعمل باستخدام بنياتContent-based ، مثل NDN، بينما يمكن للشبكة الأساسية (Core Network)أن تعمل بنموذج اتصال عالي الأداء مُعتمد على العناوين (IP-

Based) وتم استخدام SDN Controller لتوجيه الرُزم ضمن ال NDN Fabric. تم التوصية ضمن هذا البحث بتطبيق NDN–Fabric على مجال أوسع وبالتالي اعتبرنا هذا البحث دافع وخطوة باتجاه تطوير SDN–like كمستوى تحكّم لل NDN Data plane على كل طبقات الشبكة.

قامت الدِراسة [11] بتقييم فعالية شبكات البيانات المُسمّاة (NDN) في سياق شبكات الحافة التكتيكيّة (tactical edge networks) وإمكانية اعتمادها. وضّح الباحثون هنا أنّه على الرغم من أنّ بُنية الشبكات ICN تُقدّم بعض الفوائد الهيكلية للبيئات التكتيكية (المتحرّكة) أكثر من بُنية شبكات TCP / IP، إلا أن عددًا من التحديات بما في ذلك التسمية والأمن وضبط الأداء وما إلى ذلك، لا تزال بحاجة إلى المعالجة من أجل التبنّي العملي . كما قاموا بشرح للميزات التي تجعل NDN مؤهلة للتطبيق ضمن الشبكات المُتحرّكة (الشبكات التكتيكيّة العسكريّة). يمكن أن نقوم بدراسة وفهم هذه التحدّيات والميزات كي تكون كتوصيف ضمن أي تصميم مُحتمل عند التكامل مع شبكات SDN.

بينما قد تم ضمن الدراسة [5] استكشاف تكامل شبكات SDN و NDN، وتحديد مجموعة من التحديات وطرح العديد من أسئلة البحث. يتمثل أحد الشواغل الرئيسية في قابلية توسع نماذج SD–NDN، حيث تم اختبار معظم النماذج الحالية ضمن نطاقات شبكات محدودة فقط. علاوة على ذلك، يُمثل التوافق بين وحدات تحكم SDN الحالية وإطار عمل NDN مشكلة كبيرة، مما يستلزم تطوير برامج تحكم متخصصة شبيهة بـ SDN ومصممة خصيصًا لشبكات NDN. توضح البُنية التي قدمناها نقطة رئيسية، وهي أن غالبية بنيات SD–NDN قد تُقوّض، عن غير قصد، الفوائد التي تهدف NDN إلى توفيرها، مما قد يجعل نموذج SD–NDN أقل فعالية من نموذج NDN خالص أو نموذج مُركّز على المضيف. سنقوم باعتماد جزء من البيئة المُقترحة في هذه الورقة ضمن البُنية SD–NDN وحيدة المُتحكّم، كما سوف نتبناها ضمن البنية المُوزَعة الخاصة بنا.

قام الباحثون في الورقة البحثيّة [12] بدراسة بنية الشبكات المُعرّفة برمجيّاً وإثبات الفائدة من الانتقال إلى هذا النوع من الشبكات عن طريق مقارنة قيم معاملات الأداء بين هذه الشبكات والشبكات التقليدية، كما قاموا بعرض لتصنيفات شبكات SDN من حيث بُنيتها

(البُنية المركزيّة والبُنية المُوزّعة) بالإضافة لمُقارنة بين هذه البُنى لتحديد البُنى الأفضل والتي من خلالها يمكن تحقيق أداء أفضل وإتاحية أعلى. اعتمدنا على هذه الدِراسة في فهم نظريّ وعمليّ للبُنية المركزيّة (وحيدة المُتحكّم) مما سهّل عملية التحقيق لدينا أكثر.

قدّم الباحثون في الورقة البحثيّة [13] دِراسة لمِعمارية الشبكات NDN وطريقة عملها وآلية تقييم الأداء باستخدام المُحاكي ndnSIM، كما عملوا على دِراسة منهجية إعادة توجيه رُزم الاهتمام (Interest) واستراتيجيات إعادة التوجيه المختلفة في الشبكات NDN بقصد تحديد خصائص كل استراتيجية مع التركيز على اختبار مستوى توافقها مع سياسة استبدال ذاكرة التخزين المؤقت

Improved Cache replacement policy based-on Content Popularity (ICCP) المُرتبطة بمفهوم ال Cashing. ساعدتنا هذه الدِراسة في فهم والتعرّف على معمارية شبكات NDN آلية تحقيقها واختبارها ضمن المُحاكي ndnSIM.

رأينا في الدراسات السابقة دراسات للبُنى SDN و NDN كُلّاً على حِدى كما رأينا عدة نماذج وطوبولوجيا مُقترحة للبنية الجديدة SD-NDN مع عدم وجود وسطاء مُحاكاة مُناسبة لبحثنا لكن تم إيجاد الحل في القسم 1.8. لقد قمنا ضمن كل دراسة بتبيان دورها ومدى استخدامها في بحثنا وكان أهمها اعتماد البُنية المُقترحة ضمن الورقة البحثيّة [5].

حقّقنا في دراستنا البنية وحيدة المُتحكّم وبعدها قمنا بتوسعتها لتشمل البنية المُوزّعة ضمن البيئتين IP-based و NDN-Based، كما وقمنا بالمقارنة بين أهم مُعاملات الأداء من إنتاجية وتأخير وفقدان الرُزم واستهلاك عرض الحزمة بصدد إظهار التحسينات التي أضافتها شبكات NDN ضمن بنيتنا الجديدة SD-NDN وحيدة المُتحكّم.

٨. تحقيق البُنية SD-NDN وحيدة المُتحكّم في المُحاكيين Mininet و ndnSIM:

سيتم بعد تحديد وسطاء التجارب، تحقيق البُنيتين وحيدة المُتحكّم ضمن البيئتين IP-Based (على المُحاكي Mininet) والبيئة NDN-Based (على المُحاكي ndnSIM).

١,٨ وسطاء التجارب

لم نجِد ضمن الدِراسات ما هو واضح من وُسطاء لتطبيق المُحاكاة، فقُمنا بالاستِعانة بوسطاء وِفقاً لقيم نموذجيّة مأخوذة من شبكات البنية الجديدة SDN-Based Internet2 (+2014)مع شبكات NDN.

إنّ Internet 2 هي عِبارة عن شبكة تعاونية للأبحاث والتعليم تم تطويرها بواسِطة اتحاد Internet2 consortium، وهي مُنظمة غير ربحية مقرُّها الولايات المُتحدة الأمريكيّة تعمل على تشغيل تقنيات الشبكات المتقدمة للمؤسسات الأكاديمية والبحثية.

يُوضّح الجدول (١) القيم النموذجية لوسطاء المُحاكاة فيما يتعلّق بشبكات NDN:

جدول (١) القيم النموذجية لوسطاء المُحاكاة وِفق Internet 2

Metric	Description	Typical Values
Simulation Time	The total duration of the simulation.	100s - 1000s (depends on the network size and objectives)
Average Latency (Delay)	The round-trip time (RTT) for packets to travel from consumer to producer and back.	30ms - 180ms (depends on network scale and path length)
Queue Size	The maximum number of packets that can be stored in a router's queue before dropping.	50 - 200 packets (varies with network load and traffic)
Data Rate	The rate at which data is transmitted over the network.	100 Mbps - 10 Gbps (depending on link speed and capacity)
Content Store Size	The maximum number of entries (packets/data chunks) that can be stored in the Content Store (CS).	1000 - 5000 entries
Interest Request Frequency	The frequency of Interest packets sent by consumers to request data.	1 per second to 10 per second (depends on consumer behavior and simulation setup)

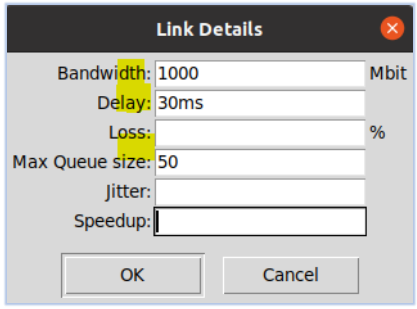
بينما نُوضّح بالجدول (٢) الوسطاء التي سنستخدمها ضمن سيناريو البنية SD-NDN بمُتحكّم وحيد (Single Controller):

جدول (2) وسطاء مُحاكاة البنية SD-NDN Single Controller

الوسطاء	القيمة
زمن المُحاكاة	100 (Sec)
عدد المتحكمات	1
عدد العقد NDN	3 (Mesh)
عدد أزواج ال Producers/Consumers	1
حجم البيانات المُرسلة (Payload Size)	10 (MB)
التأخير على الوصلات (Delay)	30 (ms)
معدّل نقل البيانات (Data Rate)	1 (Gbps)
حجم الرتل على منافذ العُقد (Queue Size)	50 (packets)
حجم ذاكرة تخزين المحتوى (Content Store)	1000 (entries)
تردد الطلبات (Interest Request Frequency)	10 (pkt per second)

عدد تكرارات المُحاكاة	5

إن تحديد هذه الوسطاء ضمن المُحاكي Mininet لا يتطلّب مجهود كبير بل يتم عبر واجهة المُحاكي: (الشكل (٣))



الشكل (٣) تحديد وسطاء المُحاكاة ضمن المُحاكي Mininet

أما ضِمن المُحاكي ndnSIM لا بد من تعديل قيم هذه الوسطاء ضمن الشيفرة المصدريّة وِفق التالي (الشكل (٤))

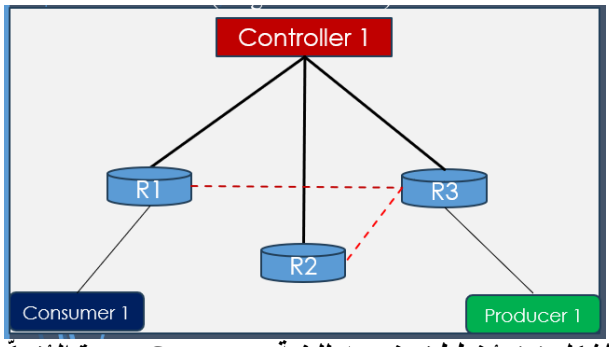
الشكل (٤) تحديد وسطاء المُحاكاة ضمن المُحاكي ndnSIM

قُمنا بتحقيق الشيفرة المصدريّة وبناء الملف الخاص بالطوبولوجيا الشبكيّة :

https://github.com/Ali-Mustafa96/Single-SD-NDN.git

٢,٨ تحقيق البنية SD-NDN وحيدة المُتحكّم

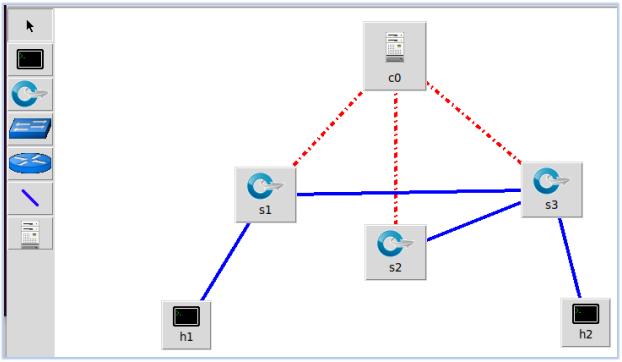
يُوضّح الشكل (٥) مُخطط البُنية وحيدة المُتحكّم التي سنعتمدها، والتي ستتضمّن مُتحكّم وحيد (Controller 1) وثلاث عُقد NDN (R1, R2, R3) وعقدة المُستهلِك المُرسل للطلبات (Consumer 1) وعقدة المُنتج المُستقبِل للطلبات (Producer 1):



الشكل (٥) مُخطط توضيحيّ للبنية SD-NDN وحيدة المُتحكّم

تم تحقيق هذه البنية ضمن المُحاكي Mininet (بيئة IP–Based) وتشغيل الطوبولوجيا وإرسال حجوم بيانات بين عقدتي الإرسال والاستقبال وِفقاً للحجم المُتفق عليه ضمن وسطاء المُحاكاة.

نُوضّح بالشكل (٦) شكل الطوبولوجيا وحيدة المُتحكّم الناتجة ضمن المُحاكي Mininet.



الشكل (٦) تحقيق البنية SDN IP-Based ضمن المُحاكي Mininet

ثمّ تم تحقيق هذه البنية ضمن المُحاكي ndnSIM (بيئة NDN–Based) وتشغيل الشيفرة المصدريّة وإظهار الطوبولوجيا ومن ثُمّ إرسال حُجوم بيانات بين عُقدتي الإرسال والاستقبال وِفقاً للحجم المُتفق عليه ضِمن وسطاء المُحاكاة.

نُوضّح بالشكل (7) الملف النصيّ الخاص بالطوبولوجيا والذي ستتم قراءتهُ ضمن الشيفرة المصدريّة الخاصة بإعدادات المُحاكاة. يتكوّن هذا الملف من قِسمين رئيسيين، الأول لبِناء العُقد وتحديد مواقِعُها (وِفق مِحوريّ x و y) والقِسم الثاني لإنشاء الوصلات وتحديد خصائصها من سرعة وتأخيرات وحجم الأرتال.

```
**/
router
#node   city    y   x   mpi-partition
leaf-1  NA  -80 -40 1
leaf-2  NA  -80 20  2
rtr-3   NA  -70 -10 0
rtr-1   NA  -60 -20 1
rtr-2   NA  -60 0   2
Cont1   NA  -40 -10 0

link

# from      to      capacity    metric  delay   queue
leaf-1      rtr-1   1Gbps       1   30ms    1000
leaf-2      rtr-2   1Gbps       1   30ms    1000
rtr-3       rtr-2   1Gbps       1   30ms    1000
rtr-3       Cont1   1Gbps       1   30ms    1000
rtr-1       rtr-2   1Gbps       1   30ms    1000
rtr-1       Cont1   1Gbps       1   30ms    1000
rtr-2       rtr-3   1Gbps       1   30ms    1000
rtr-2       rtr-1   1Gbps       1   30ms    1000
rtr-2       Cont1   1Gbps       1   30ms    1000
```

الشكل (٧) بُنية الملف النصيّ الخاص بالطوبولوجيا ضمن المُحاكي ndnSIM

نعرُض بالشكل (٨) كيفيّة إضافة عُقد التحكُّم (لدينا عُقدة مُتحكّم واحِدة في السيناريو المُوحّد المُوضّحة ضمن القسم 1.8).

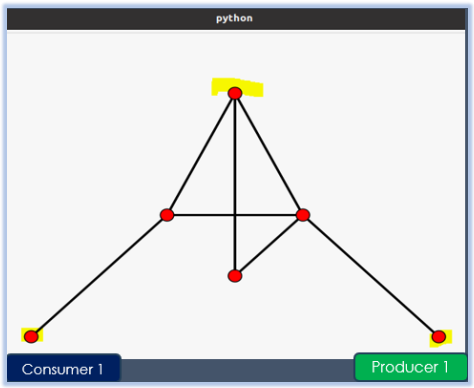
حيث أنشأنا غرض من الموديول الفرعيّ "LearningController" (الخاص ببناء عُقد المُتحكّم) المُستورد من الموديول الأساسيّ "OpenFlow" وربطنا هذا الغرض مع العُقدة المُمثّلة للمُتحكّم والتي قُمنا بقراءتها من ملف الطوبولوجيا مع تحديد ال Container الحاوية للعّقد المّتبقيّة (Consumer, Producer and NDN nodes) بالشكل المُناسب.

```
// Creating nodes  from topology text file
AnnotatedTopologyReader topologyReader("", 1);
topologyReader.SetFileName("src/ndnSIM/examples/topologies/topo-tree-Single.txt");
topologyReader.Read();

   // Getting containers for the consumer/producer and controller Cont1
Ptr<Node> consumer = Names::Find<Node>("leaf-1");
Ptr<Node> producer = Names::Find<Node>("leaf-2");
Ptr<Node> switchNode = Names::Find<Node>("Cont1"); // The controller node
```

```
//Ptr<Node> switchNode = SwitchContainer.Get (0);
OpenFlowSwitchHelper swtch;

if (use_drop)
  {
    Ptr<ns3::ofi::DropController> controller = CreateObject<ns3::ofi::DropController> ();
    swtch.Install (switchNode, switchDevices, controller);
  }
else
  {
    Ptr<ns3::ofi::LearningController> controller = CreateObject<ns3::ofi::LearningController> ();
    if (!timeout.IsZero ()) controller->SetAttribute ("ExpirationTime", TimeValue (timeout));
    swtch.Install (switchNode, switchDevices, controller);
  }
```

الشكل (٨) إضافة عُقدة المُتحكّم ضِمن الشيفرة المصدريّة في المُحاكي ndnSIM

نُوضّح بالشكل (٩) شكل الطوبولوجيا وحيدة المُتحكّم الناتجة ضمن المُحاكي ndnSIM.



الشكل (٩) تحقيق البنية SDN NDN-Based ضمن المُحاكي ndnSIM.

٩. اختبار أداء البنيتين وحيدة المُتحكّم ضمن البيئتين IP–Based و NDN–Based وتحديد التحسينات في البنية الجديدة:

سيتم ضمن كل السيناريوهات اختبار الأداء بما يتعلّق بالإنتاجية والتأخير وفقدان الرُزم واستهلاك عرض الحزمة. (حصلنا على نتائج مُقارِبة مع تكرار عملية المُحاكاة عدة مرات)

1.9 اختبار أداء البُنيتين وحيدة المُتحكّم ضمن البيئتين IP-Based و NDN-Based:

– من ناحية الإنتاجية (Throughput):

إنّ الإنتاجية Throughput تُعبّر عن كمية البيانات المُستلمة بشكل صحيح خلال زمن مُحدّد.

ضمن البيئة IP-Based: لقد فرضنا ضمن وسطاء المُحاكاة أن حجم حِمل البيانات يبلغ ١٠ MB وتساوي ١٠،٢٤٠ KB (10,240,000 Bytes) أي ما يُعادل ١٥٨ رسالة ICMP بحجم ٦٥،٠٠٠ بايت للرسالة. نتج لدينا Throughput بمقدار 0.473 Mbps

بينما ضمن البيئة NDN-Based: بسبب ندرة الرُزم المفقودة بلغت الإنتاجية 0.8 Mbps.

- **من ناحية التأخير الحاصل على إيصال الرُزم (Delays):**

ضمن البيئة IP-Based: كان أدنى تأخير 128.773 ms (ميلي ثانية) بين الرُزم وأعظم تأخير ٥٩٧,٤٨٧ ms ، بينما بلغ مُتوسّط التأخير لجميع الرُزم: ١٩٩,٥٧٨ ms.

بينما ضمن البيئة NDN-Based: كان التأخير أفضل وأقل وبلغ ١٨٠,٢٥٣ ms.

- **من ناحية فُقدان الرُزم (Dropped Packets):**

ضمن بيئة IP-Based: اكتشفنا وجود 14 رُزم مفقودة. تحتاج الموجهات لبعض الزمن كي تستقبل أوامرها من المُتحكّم مما يسبب ضياع بعض الرُزم بالإضافة لأسباب أخرى سنحدّدها ضمن قسم **تحديد التحسينات التي قامت بها شبكات NDN في البنية الجديدة.**

ضمن بيئة NDN-Based: لم نلحظ أي فقدان للرُزم وسبب ذلك خاصية التخزين المؤقت ضمن شبكات NDN وبالتالي القدرة على تلبية الطلبات Interest دون حصول فقد.

- **من ناحية استهلاك عرض الحِزمة المُستخدَم (Bandwidth):**

ضمن بيئة IP-Based:

قمنا بتحديد عرض الحزمة الكليّ 1 Gbps على جميع الوصلات، لكن تستهلك SDN في السيناريوهات الصغيرة من 1 % (أو أقل حسب حجم الشبكة) إلى 5 % من عرض ال Bandwidth الكليّ. وقد بلغ مقدار عرض الحزمة المستخدم 0.52 Mbps.

ضمن بيئة NDN-Based:

بما أنه لا يوجد رُزم Dropped، تكون جميع الرُزم المرسلة مُستلمة ويكون عرض الحُزمة المُستهلك هو حجم البيانات المُرسلة (Payload) خلال زمن المُحاكاة (100 sec) وقد بلغ0.8 Mbps .

وضعنا ضِمن الجدول (٣) مُجمل النتائج التي حصلناه عليها من ناحية التأخير وخسارة الرُزم والإنتاجية:

جدول (3) مُجمل نتائج المُحاكاة من تأخير وخسارة رُزم وإنتاجية

Architecture	AVG Delay(ms)	DROP number	Throughput (Mbps)
IP-Based (Single Controller)	199.578	14	0.473
NDN-Based (Single Controller)	180.253	0.4	0.8
IP-Based (Multi-Controller)	330.141	9	0.486
NDN-Based (Multiple Controller)	300.421	0.4	0.8

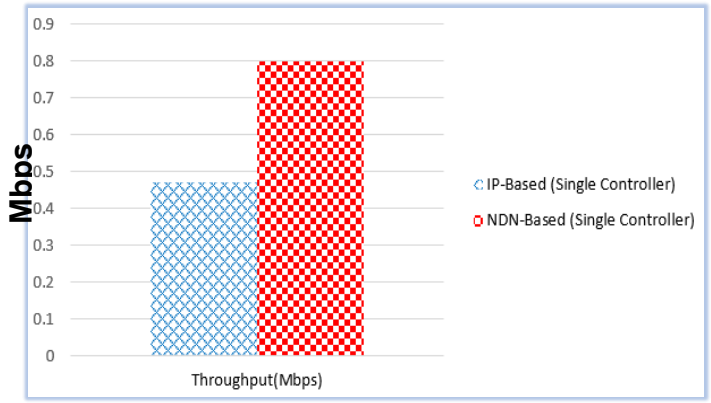
كما نُوضّح بالجدول (٤) قيم استهلاك عرض الحزمة:

جدول (4) الفرق باستهلاك عرض الحزمة بين البيئتين IP-Based و NDN-Based

Architecture	Bandwidth(Mbps)
IP-Based Architectures	0.52
NDN-based Architectures	0.8

2.9 تحديد التحسينات التي قامت بها شبكات NDN في البنية الجديدة:

إن التحسُّن في الإنتاجية كما في الشكل (١٠) سببه قِلة الرِّزم المفقودة في البيئة NDN–Based وسنشرح الأسباب عند استعراض مُخطّطات فقدان الرُزم.



الشكل (١٠) مُخطط يُوضّح الفرق بالإنتاجية ضمن البُنيتين IP-Based Single Controller

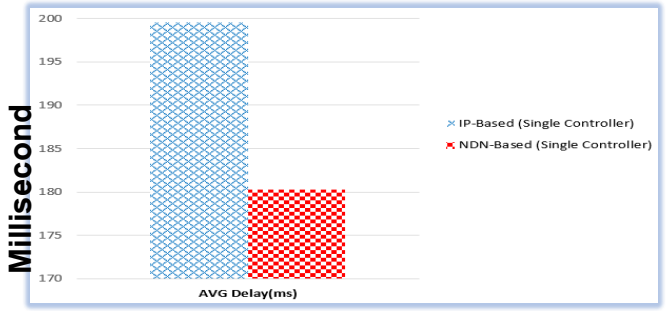
نعرِض فيما يلي أسباب التحسينات التي أضافتها شبكات NDN ضمن البنية SD–NDN فيما يتعلّق بالتأخير الحاصِل كما في الشكل (١١):

١– التوجيه المُتكيّف (Adaptive Forwarding):

تدعم شبكة NDN إعادة التوجيه متعدد المسارات وفي حال ازدحام الرابط، تُعيد شبكة NDNتوجيه حركة المرور تلقائيًا عبر مسارات بديلة) تُقلل شبكة NDN من التأخيرات من خلال اختيار أفضل مسار ديناميكيًا دون انتظار تحديثات وحدة تحكّم (SDN) بينما تعتمد شبكة SDN القائمة على IP عادةً على مسارات مُحددة مسبقًا، وتتطلب إعادة التوجيه تدخل وحدة التحكم، مما يُسبب تأخيرات. كما أنّهُ **لا حاجة للاتصال نهاية إلى نهاية (End–to–End Connectivity)**، حيثُ تقوم شبكات NDN بتوجيه البيانات بناءً على أسماء المحتوى بدلاً من عناوين IP، مما يُلغي الحاجة إلى اتصالات End–to–End. تتيح هذه الخاصية استرجاع البيانات من أي عقدة تُخزّن المحتوى مؤقتًا، مما يُقلل من التأخيرات المرتبطة بإعدادات الاتصال. [14]

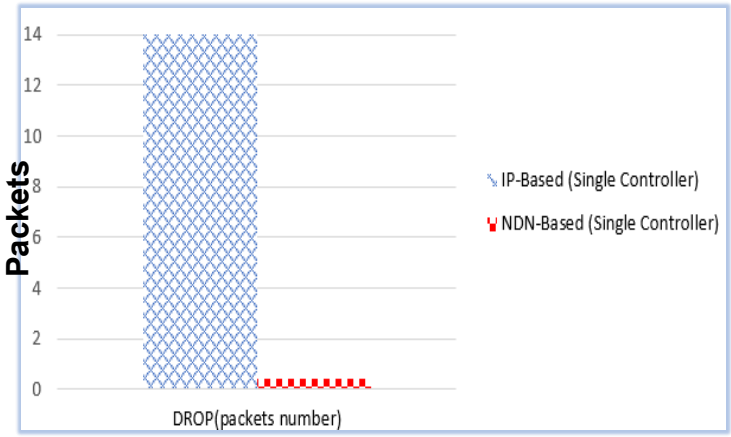
٢- لا يوجد عِبء TCP/IP:

يؤدي التخلّص من تكلفة TCP/IP ضمن بيئة NDN إلى تقليل تأخيرات المعالجة وتحسين أداء الشبكة. لا تزال شبكة SDN القائمة على IP تستخدم بروتوكول TCP/IP، مما يتطلب تأكيدات ACK وإعادة الإرسال ومُعالجة الأخطاء، مما يزيد من التأخير. [15]



الشكل (11) مُخطط يُوضّح الفرق بالتأخير ضمن البُنيتين Single Controller

وكما نرى التحسُّن الملحوظ من ناحية فُقدان الرُزم كما في الشكل (١٢) والذي أدّى بدورهِ إلى التحسّن في الإنتاجية. ونعزو ذلك إلى خاصي **التخزين المؤقت داخل الشبكة –In Network Caching)**، حيثُ يُتيح التخزين المؤقت داخل شبكات NDN والمعروفة أيضاً باسمContent-Centric Network (CCN) استرجاع البيانات من العقد الوسيطة، مما يقلل الحاجة إلى إعادة الإرسال من المصدر الأصليّ، وبالتالي تقليل فقدان الرُزم. بينما يؤدي فقدان الرُزم في شبكات SDN القائمة على بروتوكول الإنترنت (IP)، إلى إعادة الإرسال من المصدر الأصلي، مما يزيد من الازدحام والتأخير. كما أنّ نقل البيانات فقط عند الطلب عبر رُزم ال Interest، مما يساعد في التحكم في الازدحام وتقليل نقل البيانات غير الضروري، مما يؤدي إلى تقليل فقدان الرُزم. [15]



الشكل (١٢) مُخطط يُوضّح الفرق بفقدان الرُزم ضمن البُنيتين Single Controller

لكن البُنية SD-NDN ليست دائمًا أفضل من SDN المستند إلى IP من حيث استهلاك النطاق الترددي كما نلاحظ ضمن الشكل (١٣)، للأسباب التالية:

١- تستهلك حركة مرور مزامنة ذاكرة التخزين المُؤقت (Cache Synchronization Traffic) نطاقًا ترديديًا إضافيًا:

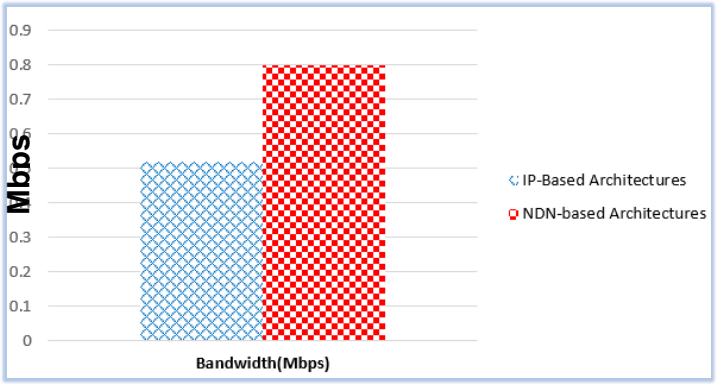
تُولّد آليات مزامنة ذاكرة التخزين المؤقت في شبكات NDN (للحفاظ على الاتساق بين أجهزة التوجيه) بيانات تحكم إضافية. بينما في شبكات IP-Based SDN، لا تحتاج أجهزة التوجيه إلى مزامنة ذاكرة التخزين المؤقت نظرًا لعدم وجود تخزين مؤقت داخل الشبكة.

باختصار: تستهلك تحديثات ذاكرة التخزين المؤقت في شبكة NDN نطاقًا ترديديًا إضافيًا، خاصةً في الشبكات عالية الديناميكية. [16]

٢- تزيد بادئات الأسماء الكبيرة من حجم الرُزمة(Large Name Prefixes) :

يستخدم NDN أسماء المحتوى بدلاً من عناوين IP، مما يعني أنه ممكن أن تحتوي ترويسات الرُزم على أسماء هرمية طويلة. بينما إنّ العنونة تكون ذات حجم ثابت في الأنظمة القائمة على بروتوكول الإنترنت (IP-Based). [16]

إذاً تستهلك البنية SD–NDN نِطاقًا ترددِيًا أكبر بسبب زيادة تكلفة رُزم البيانات، والطلبات المتكررة، ورؤوس البيانات الأكبر، وبيانات مزامنة ذاكرة التخزين المؤقت. بينما تتميّز IP–Based SDN بكفاءة أعلى في استخدام النطاق الترددي بفضل دعم البث المتعدد، ورؤوس رُزم البيانات المُدمجة، وانخفاض تكلفة التحكّم.



الشكل (13) مُخطط يُوضّح الفرق باستهلاك عرض الحزمة بين البيئتين IP-Based و-NDN Based

لكن على أية حال تمت التوصية ضمن شبكات البنية الجديدة SDN–Based Internet2 (٢٠١٤+) مع شبكات NDN بأن يكون عرض الحزمة كافي ومن مرتبة ١ Gbps وما فوق، وبالتالي فإنّ استهلاك عرض الحزمة لن يكون عائق أو تحدٍ كبير عند تحقيق الاندماج بين البُنيتين.

١٠. الخاتِمة والتطلُّعات المُستقبليّة

قدّمنا في هذه الورقة البحثيّة نموذج وتحقيق فعليّ للبنية الجديدة SD-NDN وطرحنا ما يناسبها من وسطاء مُحاكاة، وقد شمل هذا التحقيق البُنية المركزيّة (وحيدة المُتحكّم) ضمن كِلا البيئتين IP-Based التقليديّة و NDN-Based، كما حدّدنا جميع التحسينات المُمكنة التي سبّبتها البُنية NDN ضمن هذه البُنية الجديدة وقمنا بقياس أهم وسطاء الأداء ووضع التفسيرات المُناسبة لكل تحسّن تم إيجادُه.

نتطلّع الآن إلى تطوير الشيفرة المصدريّة الخاصة بالبُنية SD–NDN وحيدة المُتحكّم وتوسعتها لتُصبِح داعِمة للبَنية المُوزّعة بالإضافة لإمكانية تخصيص خوارزميات مُوازنة

الحِمل بين المُتحكّمات وتحديد أي وسطاء تتعلّق بذلك. يُمكن أن تتم مجالات أخرى على هذا البحث هي تطبيق البيئة SD-NDN على بنية مُوزّعة تشابكيّة هرميّة، ودِراسة الأداء الناتج عن هذا التكامل والتوزيع.

جدول المُختصرات:

CCN	Content-Centric Network
CS	Content Store
ICCP	Improved Cache replacement policy based-on Content Popularity
ICN	Information-Centric Network
NDN	Named Data Network
SDN	Software-Defined Network
SD-NDN	Software-Defined Named Data Network

١١. المراجِع:

[١] S. S. M.Micheal و .. Dr.K، "A SURVEY ON NAMED DATA NETWORKING."، *IEEE SPONSORED 2ND INTERNATIONAL CONFERENCE ON ELETRONICS AND COMMUNICATION SYSTEM (ICECS)*، ٢٠١٥.

[٢] A. Marica و &.others، "Understanding Name-based Forwarding Rules in Software-Defined Named Data

Authorized licensed use limited to: University of "،.Networking *New South Wales. Downloaded on September 26,2020 at 18:20:29 UTC from IEEE Xplore*، ٢٠٢٠.

[٣] "A Survey of Software-Defined ،others& و N. Bruno Networking: Past,Present, and Future of Programmable *HAL Id: hal-00825087 https://hal.inria.fr/hal-* "،.Networks *00825087v5*، ١٩ January 2024.

[٤] "A review of SDN-enabled ،A. Jamali و N. Naja ،Y. Sembati *Engineering* "،.routing protocols for Named Data Networking *.Sons Ltd & Reports published by John Wiley*، ٦ March 2024.

[٥] "Software-Defined Named Data ،B. Abdelfettah و A. Albatool *Future Internet 2024, 16,* "،Networking in Literature: A Review *258.*، ٢٣ July 2024.

[٦] L. Bob و others& ، Introduction to Mininet . ، ٢٠١٧.

[٧] Development of a performance measurement tool ،M. Marion for SDN، ٢٠١٥.

[٨] "NDN"، [متصل]. Available: https://named-data.net/.

[٩] "A Named Data Networking (NDN), ،others& و J. Zain ul Abidin New Approach to Future Internet Architecture Design: A *Communication & International Journal of Informatics* "،.Survey *Technology (IJ-ICT) Vol.2, No.3*، December 2013.

[١٠] "NDN Fabric: Where the Software- ،others& و L. R. M. André *IEEE* "،.Defined Networking Meets the Content-Centric Model

TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, VOL. 18, NO. 1، March 2021.

[١١] C. Lorenzo و &others، "Considerations on the Adoption of Named Data Networking (NDN) in Tactical Environments."، *International Conference on Military Communications and Information Systems (ICMCIS)*، ٢٠١٩.

[١٢] د. م. عبود و م. م. الضاهر، "تحليل الأداء في الشبكات المعرفة برمجياً،" *مجلة جامعة حمص المجلد ٤١*، ٢٠١٩.

[١٣] د. ي. غازي و م. ع. الحمصيه، "تقييم أداء السياسة المُحسّنة لاستبدال ذاكرة التخزين المؤقت بالاعتماد على شعبية المُحتوى (ICCP) عند استراتيجيات إعادة التوجيه المختلفة في شبكات البيانات المُسمّاة،" *مجلة جامعة حمص المجلد ٤٦ العدد ٧*، ٢٠٢٤.

[١٤] K. Sarantis و &others.، "Logically-Centralized SDN-Based NDN Strategies for Wireless Mesh Smart-City Networks."، *Future Internet 2023, 15, 19. https://doi.org/10.3390/fi15010019*، ٢٩ December 2022.

[١٥] R. S. Nana و &others.، "Performance Comparison of Named Data and IP-based Network—Case Study on the Indonesia Higher Education Network."، *Journal of Communications Vol. 13, No. 10*، October 2018.

[١٦] A. Manar و &others، "Leveraging ICN and SDN for Future Internet Architecture: A Survey."، *Electronics 2023, 12, 1723. https://doi.org/10.3390/electronics12071723*، ٤ April 2023.

تكامُل وتحسين الشبكات المُعرّفة برمجيّاً المُوزّعة باستخدام بُنية شبكات البيانات المُسمّاة

د.م. يمان غازي[١] م. علي مصطفى[٢]

المُلخّص

لقد تناولنا في هذا البحث تحقيق التكامُل بين الشبكات SDN والشبكات NDN للحصول على البُنية الجديدة SD-NDN في المُحاكي ndnSIM، وقُمنا هنا بالتخصيص وتبني البُنية المُوزّعة وذلك بكتابة الترميز المصدريّ الذي يُحاكي عملَها بالإضافة إلى تضمينهِ في المحاكي، كما قُمنا بتحقيق البُنية IP-Based المُشابهة لها ضمن المحاكي Mininet من أجل المُقارنة بين البنيتين في البيئة NDN-Based والبيئة IP-Based حيث سوف يتم اسقاط ذلك من خلال إنشاء طوبولوجيا مُوحّدة وتحديد وسطاء تجربة مُناسبة، ومن ثم بدأنا بالتحقيق من خِلال تنفيذ سيناريو عمليّ على كلا المُحاكيين، وفي الجزء الأخير من هذا البحث تمَّ تقييم أداء البُنى المُحقّقة ومُقارنة الشبكة IP-Based مع شبيهتِها ضِمن البيئة NDN-Based. أثبتنا بالنتائج تفوّق البُنية ضمن بيئة NDN من ناحية الإنتاجية وتقليل الرُزم المفقودة والتأخير مُقارنةً مع البيئة التقليدية IP-Based، بينما ظهر لدينا أنَّ استخدام عرض الحزمة كان أكبر في البُنية NDN-Based.

الكلمات المفتاحية: شبكات البيانات المُسمّاة، الشبكات المُعرّفة برمجيّاً، الشبكات المُعرّفة برمجيّاً المُوزّعة، المُحاكي ndnSIM، المُحاكي Mininet، الشبكات المركزيّة، إنترنت المستقبل، شبكات SDN NDN (SD-NDN).

[١] استاذ مُساعد، قسم هندسة الشبكات والنظم الحاسوبية، كلية الهندسة المعلوماتية، جامعة حمص
[٢] طالب دراسات عليا (ماجستير)، قسم هندسة الشبكات والنظم الحاسوبية، كلية الهندسة المعلوماتية، جامعة حمص

Integrating and Improving Distributed Software–Defined Networks Using Named Data Networks Architecture

Dr. Eng. Yaman Ghazi[1] Eng. Ali Mustafa[2]

ABSTRACT

In this research, we have addressed the integration of SDN and NDN networks to obtain the new SD-NDN architecture in the ndnSIM simulator. Here, we have customized and adopted the distributed architecture by writing the source code that simulates its operation and including it in the simulator. We have also implemented the similar IP-Based architecture in the Mininet simulator to compare the two architectures in the NDN-Based environment and the IP-Based environment. This will be projected by creating a unified topology and determining appropriate experiment brokers. Then, we started the investigation by implementing a practical scenario on both simulators. In the final part of this research, we evaluated the performance of the implemented architectures and compared the IP-Based network with its counterpart in the NDN-Based environment. We demonstrated that the NDN architecture outperforms the traditional IP-based architecture in terms of throughput, packet loss reduction, and latency, while bandwidth utilization was higher in the NDN-based architecture.

[1] **Assistant Professor, Department of Network Engineering and Computer Systems, Informatics Engineering Faculty, Homs University.**

[2] **Postgraduate Student (M.A), Department of Network Engineering and Computer Systems, Informatics Engineering Faculty, Homs University.**

Keywords: Named Data Networking, Software Defined Networking, Distributed Software Define Networking, Information Centric Networking, Future Internet, SDN NDN Networks (SD-NDN).

١- مُقدمة:

إنَّ شبكات البيانات المُسمّاة (NDN) هي بُنية شبكات مركزيّة للمعلومات (ICN) تُحوّل نموذج الإنترنت من مركزيّة المضيف إلى مركزيّة المعلومات، مع استخدام أسماء المحتوى على مستوى التطبيق مباشرةً في طبقة الشبكة للبحث عن البيانات واسترجاعها. تعتمد شبكات NDN على تبادل نوعين من الرُزم: رُزم الاهتمام (Interest)، الذي يستخدمها المُستهلك (Consumer) لطلب المحتوى، ورُزم البيانات (Data)، التي يستخدمها المُزوّد (Producer) للرد على الطلب. بالمقارنة مع تصميم IP، فإن مستوى بيانات NDN هو حالة، ومتكيف، وينفذ التخزين المؤقت داخل الشبكة. عندما تستقبل عقدة NDN حزمة اهتمام، فإنها تبحث أولاً عن نتيجة بيانات مخزنة مؤقتًا في مخزن المحتوى المحليّ (Content Store - CS). في حالة الفشل، تبحث عن اسم مطابق في جدول الاهتمامات المُعلّقة (PIT)، الذي يسجل الاهتمامات المرسلة التي لم يتم تلبيتها بواسطة البيانات بعد. إذا تم العثور على تطابق، يتم تجاهل الطلب لأن هناك اهتمامًا مُساويًا معلقًا بالفِعل. إذا فشل كل من فحص المخزن CS والجدول PIT، يمكن للعقدة إعادة توجيه الرُزمة، وِفقًا للقواعد في قاعدة معلومات التوجيه (FIB)، المملوءة إما ببروتوكولات التوجيه أو من قبل مسؤولي الشبكة. **[1]**

يُحاول المُصممون ضمن شبكات SDN إعادة ترتيب أجزاء وأدوار كل مُكونات البُنية التحتيّة للشبكات (infrastructure)، والتي لم تُعدّل منذ عام 1980 حيث أن أخر تعديل كان هو الانتقال من شبكات الجيل القديم NCP إلى شبكات TCP/IP الغنيّة عن التعريف، ومنذ ذلك الحين لم يطرأ أي تغيّير على مُستوى البُنية التحتيّة للشبكات لتتلاءم والتطوُر

الكبير الحاصِل في مجال تقنية المعلومات، وخصوصاً التقنية التخيليّة virtualization التي قامت بمحاكاة افتراضية لجميع طبقات تصميم الشبكات، ولكن ظلّت طبقة البُنية التحتيّة للشبكات مُستعصية على هذه التقنية. إذاً شبكات SDN هي محاولة ناجحة لفصل مُستوى البيانات Data plane أو ما يُعرف باسم Forwarding plane عن مُستوى التحكُّم control plane، ليصبح دور أجهزة الشبكات Switches/Routers مُنحصِر فقط في تمرير البيانات forwarding، أما عمليات الإدارة والتحكُّم والخدمات ستُصبِح في طبقات جديدة وهي طبقتين مُنفصلتين.

عرَّفت مُنظمة ONF البُنية المِعماريّة لتقنية SDN على شكل نموذج (Module) مُكوّن من ثلاث طبقات، أو بمعنى أخر فإنّ SDN تقسم البنية التحتية للشبكات إلى ثلاث طبقات (Data, Control and Application). [**14**]

٢- الهدف من البحث:

ساهمنا عبر هذا البحث في تحقيق تكامُل فعليّ لبُنية شبكات NDN مع بُنية شبكات SDN باستخدام عدة مُتحكُّمات ضمن المُحاكي ndnSIM وذلك بكتابة الترميز المصدري الذي يبني ويُحاكي عمل هذه البُنية وتضمينهُ في المُحاكي، ثُمَّ استخدام المُحاكي Mininet لمُحاكاة ذات البُنية المُتكاملة. كما تتمثّل الأهداف الخاصة في تحقيق بُنية SDN مُوزّعة مع شبكات NDN وقياس مُعامِلات الأداء، وتحديد التحسينات التي سنحصل عليها عند تحقيق تكامُل هذه البُنى.

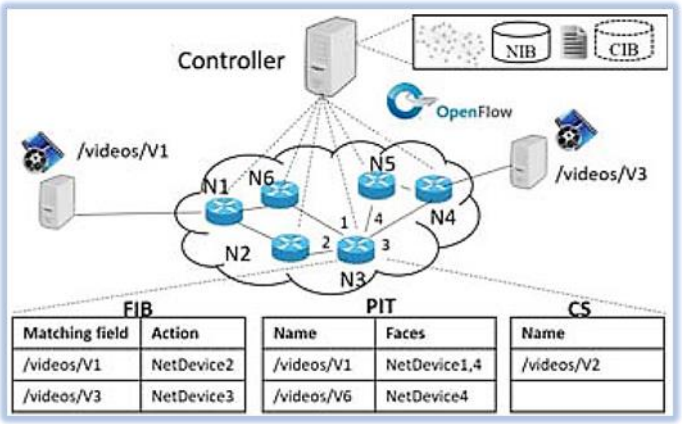
٣- البُنية SDN NDN المُتكاملة الهجينة:

يُوفّر نهج شبكات NDN مزايا عديدة مقارنةً بالنهج القائم على IP. على سبيل المثال، فهو يقلل من أعباء المعالجة ضِمن الشبكة ويُحسّن توافر المحتوى بفضل التخزين المؤقت داخل الشبكة، والذي يسمح لأي عقدة بتخزين البيانات وتقديمها. كما أنه يتيح نقل البيانات بكفاءة، بفضل التوصيل المُدار من قِبل المُستقبِل الذي يرسل رُزم البيانات فقط

استجابةً لرُزم الاهتمام (Interest) من المُستهلكين، مما يتجنّب البيانات غير المرغوب فيها أو المكررة. علاوةً على ذلك، فإنَّ أي عُقدة ضِمن الشبكة قادرة على تأمين الأمان والتشفير اللازم من تلقاء ذاتها بدون اللجوء لأطراف خارجيّة لتأمين الحماية. **[3]**

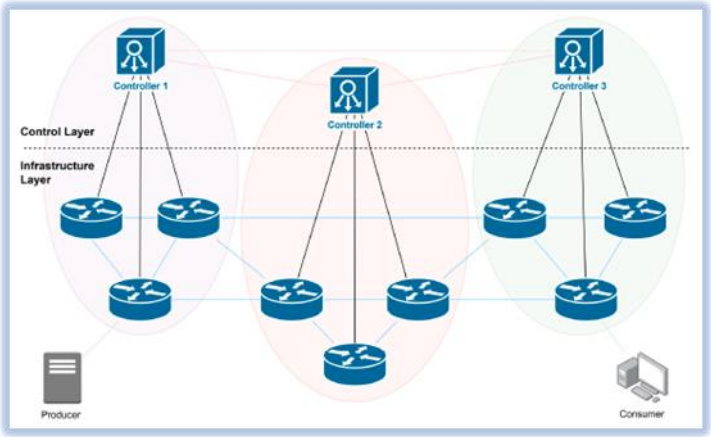
في الآوِنة الأخيرة، تم اعتبار الشبكات المُعرّفة برمجيّاً (SDN) حلاً مُحتملاً ليتم دمجه مع شبكات NDN، من أجل تحسين ميزات مستوى التحكُّم وتطبيق سياسات هندسة حركة المرور. تقوم شبكات SDN بفصل مُستوى التحكُّم عن مُستوى البيانات، ونتيجة لذلك يتم تنفيذ قرار التوجيه بواسطة كيان مركزي منطقيًا، وهو وحدة التحكُّم. في تصميم SDN NDN المُتكامُل، يتم تزويد عُقد الشبكة بجداول مستوى بيانات NDN، مع ملء FIBs بواسطة وحدة تحكُّم SDN. عندما لا يمكن التعامل مع حزمة مهمة محليًا في مستوى بيانات NDN، يتم الاتصال بوحدة تحكُّم SDN التي تنفذ قرار التوجيه وتحفُن قواعد إعادة التوجيه القائمة على الاسم في FIB وفقًا لذلك. في شبكة ذات عدد كبير من طلبات المحتوى، قد تتعرّض قواعد بيانات FIBs للتدفق الزائد بسهولة نظرًا لحجم الجدول المحدود، مع ما يصاحب ذلك من مشاكل في قابلية التوسع. في الواقع، يجب الحفاظ على حجم قاعدة بيانات FIB صغيرًا لسرعة البحث وأداء توجيه جيد. قد يؤدي التدفق الزائد في قاعدة بيانات FIB إلى حظر الطلبات الجديدة أو حذف الإدخالات الحالية لاستيعاب القواعد الجديدة التي أدخلها المتحكُّم. وبالتالي، فإن القرارات المتعلقة بالإدخالات التي يجب تخزينها في قاعدة بيانات FIB ومدة الاحتفاظ بها تؤثر بشكل كبير على أداء توجيه الرُزم. **[1]**

يُوضّح الشكل (١) البُنية المُقترحة لشبكات SD-NDN وِفق المرجِع **[1]**، وهي بُنية وحيدة المُتحكُّم وقُمنا فقط بدِراسة وفهم للبُنية انطلاقاً منها.



الشكل (١) بُنية SDN- NDN المُقترحة ضِمن المرجع [1]

يُمثل الشكل (٢) مثالاً على بُنية أُخرى لكنها مُفصّلة وواضِحة أكثر وتُقسّم هذه البُنية الشبكة إلى طبقة التحكُّم وطبقة البُنية التحتيّة (العُقد).



الشكل (٢) مِثال أوضح عن بُنية SD-NDN المُوزّعة [4]

ستكون مهمة عُقد التوجيه هنا هي إعادة توجيه الرُزم، وبالتالي تنتقل أعباء التوجيه إلى أجهزة المُتحكّمات. **[4]**

سنقوم باعتماد شكل هذه البيئة (الشكل ٢) ضمن البُنية SD-NDN المُوزَعة الخاصة بنا، وسنستخدم وسطاء خاصة بالمُحاكاة وسنقوم بذِكرها لاحقاً.

٤- أدوات تحقيق البحث

لقد استخدامنا في الاختبارات والتحقيق برنامج المُحاكاة **Mininet** (المُنصّب على التوزيعة Ubuntu 20.04 من نِظام Linux) لنمذجة وتحقيق بُنية الشبكات المُعرّفة برمجيّاً حيثُ يسمح هذا المُحاكي باختبار الطوبولوجيا بأنواعها دون الحاجة إلى شبكة مادية للتحكُّم في الشبكة الافتراضيّة، كما يتضمّن Mininet واجهة سطر أوامر (CLI) كما يُوفّر واجهة برمجيّة مُحقّقة بلُغة Python تُساعِد في بِناء الشبكة واختبار الأداء. **[6]**

كما قُمنا باستخدام المُحاكي **ndnSIM** لإجراء تجارب المُحاكاة ودِراسة النتائج ضمن البيئة NDN-Based، حيثُ يُمكّننا من إنشاء طوبولوجيا الشبكة وتحديد مُعلّماتِها ومُحاكاة الاتصال بين عُقد NDN المُختلفة وتسجيل أحداث المُحاكاة بالإضافة إلى عرض النتائج بشكل واضح.

إنّ المكاتب ndn-cxx و NFD ضِمن المُحاكي ndnSIM تضمن أن تكون المُحاكاة واقعية إلى أقصى حد. **[14]**

لقد تم إجراء تجارُب وسيناريوهات المُحاكاة باستخدام بيئة ndnSIM 2.9 (وهو الإصدار الحالي).

٥- الدِراسات السابِقة:

لقد بحثنا ضِمن الكثير من الدِراسات التي تتناول موضوع التكامُل في البُنية بين شبكات NDN وشبكات SDN فمثلاً يقوم الباحثين في الورقة البحثية **[1]** بتطبيق بُنية SDN NDNباستخدام بروتوكولOpenFlow كما هو مُستخدم في المراجع **[7]** و **[8]**، حيث تمَّ هُنا نقل عمليات الإدارة واتخاذ قرارات التوجيه إلى عُقدة المُتحكّم، وتم الحصول على

تحسينات في توجيه الطلبات. سنعتمد في هذه الورقة على دِراسة هذه البُنية واحتمال استخدامها لاحقاً ضمن أي توسعات مستقبلية مع بُنية الشبكات SDN.

بينما اقترح الباحثون في الورقة البحثيّة **[2]** بُنية هجينة تُسمّى NDN–Fab، تسمح بالاتصال بين بُنى الشبكة المُعتمدة على العناوين (Ip–based) والمُحتوى (Content–based)، مثل NDN، وتم استخدام مُتحكّم SDN لتوجيه الرُزم ضمن البُنية الهجينة الجديدة. كما تمت التوصية بتطبيق البُنية NDN–Fabric على مجال أوسع وهو ما اعتبرناه دافع باتجاه تطبيق البنية الجديدة على كل طبقات الشبكة.

ركّزت الورقة البحثيّة **[5]** على دِراسة فعالية شبكات البيانات المُسمَّاة في سياق الشبكات المُتغيرة التكتيكية (Tactical networks)، وضّح الباحثون هنا أنّه على الرغم من أن بُنية ICN تُقدّم بعض الفوائد للبيئات التكتيكية (المتحرّكة) أكثر من البُنية التقليديّة TCP / IP، إلا أن تحديات التطبيق والتوسِعة كبيرة. كما قاموا بعرض السمات التي جعلت شبكات NDN مُؤهلة للتطبيق ضِمن الشبكات التكتيكيّة العسكريّة. يمكن أن نعتمد هذه السمات كتوصيف للتصميم الخاص بالبُنية الجديدة.

وقد قدّم الباحثون ضمن الورقة **[4]** عملية دمج وتكامُل شبكات SDN و NDN، وعرضوا أبرز الصعوبات التي تتحدى البُنية الجديدة، حيث يُؤدي الاندماج بين مُتحكّمات شبكات SDN وبِّنية الشبكات NDN إلى تحدٍ كبير، مما يتطلّب تطوير برامج تحكُّم متخصصة مُتوافِقة مع كِلا شبكات SDN وشبكات NDN. إنّ غالبية بُنى SD–NDN قد تُؤدي إلى إنقاص الميّزات التي تُوفّرها شبكات NDN. قمنا باعتماد جزء من البُنية المُقترحة في هذه الورقة ضمن البُنية SD–NDN المُوزّعة (مُتعددة المُتحكُّمات) والتي سنتناولها ضِمن بحثنا.

ولقد تناول الباحثون في البحث **[15]** مزايا الانتقال من الشبكات التقليديّة إلى الشبكات المُعرّفة برمجيّاً بما تحويه من وحدات تحكُّم خارجيّة مما يُوفّر بيئة أبسط لبرمجة ولإدارة

الشبكة وتحديد سُلوكها. كما قاموا بمُقارنة أداء إثنين من المُتحكِّمات مفتوحة المصدر وهي pox و Opendaylight من حيث العوامل التالية المُؤثِّرة أداء الشبكة وهي متوسط تأخير التدفق ومتوسط مُعدّل البِتات ومتوسط مُعدّل الرُزم و jitter باستخدام الأداة–D ITG لتحديد المُتحكِّم ذو الأداء الأفضل في عدد من بُنى الشبكات الأساسيّة المُضمّنة في المُحاكي Mininet وهيsingle و linear وtree. وتمت الإشارة إلى أن وِحدة التحكُّم pox توفر سهولة أكبر من ناحية تشغيلها وفهم مكوناتها وتطويرها، ولقد قمنا ضمن بحثنا هذا باستخدام هذا المُتحكِّم (POX) ضمن البُنية المُوزّعة في البيئة –IP Based.

يهدُف الباحثون ضمن الورقة البحثيّة [9] إلى دِراسة سياسات التخزين المُؤقّت للمحتوى وسياسات استبدال المحتوى ضِمن جهاز التوجيه ضِمن شبكات NDN، كما يهدف البحث من ناحية أُخرى إلى تَحسين أَداء سياسة استبدال ال Cash بالاعتماد على شعبيّة المُحتوى (CCP)، وعملوا عبر التجارِب على تحسين عِدّة وسطاء أداء مُرتبة بمفهوم ال Cashing ضِمن مُوجّهات الشبكات NDN. استفدنا من هذ البحث في دِراسة بُنية شبكات NDN الصافية وفهم كل ما يتعلّق بها من مفاهيم أساسيّة والتخزين المؤقت وسياساتهُ وأنواع الرُزم ووسطاء المُحاكاة المُستخدمة.

رأينا من ضمن ما سبق من دِراسات عِدة نماذج وبُنى مُقترحة للبُنية الجديدة SD-NDN. سنعمل بدورنا على تحديد وسطاء المُحاكاة المُناسبة وتحقيق البُنية المُوزّعة ضمن البيئتين IP-based و NDN-Based، وقُمنا بالمُقارنة بين أهم مُعاملات الأداء من إنتاجية وتأخير وفُقدان الرُزم واستهلاك عرض الحُزمة وإظهار التحسينات التي أضافتها شبكات NDN ضمن بُنيتنا الجديدة SD-NDN المُوزّعة.

٦- تحقيق البُنية Distributed SD-NDN في المُحاكيين Mininet و ndnSIM:

سيتم في هذا الفقرة تحقيق البُنية مُتعددة المُتحكِّمات ضمن البيئتين IP-Based (على المُحاكي Mininet) والبيئة NDN-Based (على المُحاكي ndnSIM)، لكن علينا أولاً وضع مُعلّمات التجارب.

١,٦ إعداد مُعلّمات المُحاكاة

قُمنا بأخذ وُسطاء وِفقاً لقيم نموذجيّة مأخوذة من شبكات البُنية الجديدة SDN-Based Internet2 (2014+) مع شبكات NDN.

ملاحظة: إنّ المجموعة Internet 2 عِبارة عن شبكة تعاونية للأبحاث والتعليم تم تطويرها بواسطة اتحاد Internet2 consortium، وهي منظمة غير ربحية مقرُّها الولايات المُتحدّة الأمريكيّة ويركّز عملها على دعم وتشغيل الشبكات المُتقدّمة للمؤسسات الأكاديمية والبحثية.

يُوضّح الجدول (١) القيم النموذجية لوسطاء المُحاكاة فيما يتعلّق بشبكات NDN:

جدول (١) القيم النموذجية لوسطاء المحاكاة وِفق توصيات مُنظمة Internet 2

Metric	Description	Typical Value
Simulation Time	The total duration of the simulation.	**100s - 1000s** (depends on the network size and objectives)
Average Latency (Delay)	The round-trip time (RTT) for packets to travel from consumer to producer and back.	**30ms - 180ms** (depends on network scale and path length)
Queue Size	The maximum number of packets that can be stored in a router's queue before dropping.	**50 - 200 packets** (varies with network load and traffic)
Data Rate	The rate at which data is transmitted over the network.	**100 Gbps - 10 Gbps** (depending on link speed and capacity)
Content Store Size	The maximum number of entries (packets/data chunks) that can be stored in the Content Store (CS).	**1000 - 5000 entries** (based on network caching strategy)
Interest Requests Frequency	The frequency of Interest packets sent by consumers to request data.	**1 per second to 10 per second** (depends on consumer behavior and simulation setup)

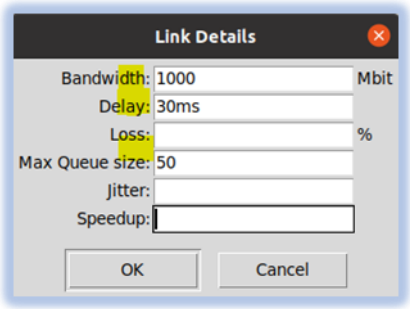
نُوضّح بالجدول (٢) الوسطاء الخاصة بالمُحاكاة ضمن سيناريو البُنية SD–NDN المُوزّعة (Distributed Controllers):

جدول(2) وسطاء مُحاكاة البُنية SD-NDN المُوزَّعة

الوسيط	القيمة
زمن المُحاكاة	100 (Sec)
عدد المُتحكِّمات	٣
عدد العقد NDN	9 (Mesh)
عدد أزواج ال Producers/Consumers	1
حجم البيانات المُرسلة (Payload Size)	10 (MB)
التأخير على الوصلات (Delay)	30 (ms)
معدّل نقل البيانات (Data Rate)	1 (Gbps)
حجم الرتل على منافذ العُقد (Queue Size)	50 (packets)

حجم ذاكرة تخزين المحتوى (Content Store)	1000 (entries)
تردد الطلبات (Interest (Request Frequency	10 (pkt per second)
عدد تكرارات المُحاكاة	5

نقوم بوضع هذه الوسطاء بنفس القيم ضمن المُحاكي Mininet، ويتم ذلك عبر واجهة المُحاكي الرسوميّة: (الشكل (٣))



الشكل (٣) تحديد وسطاء المُحاكاة ضمن المُحاكي Mininet

نضع هذه القيم في المُحاكي ndnSIM ضمن ملف الشيفرة المصدريّة الخاصة بإعدادات وطوبولوجيا الشبكة المُوزّعة وِفق التالي (الشكل (٤))

الشكل (٤) تحديد وسطاء المُحاكاة ضمن المُحاكي NDN SIM

عِلماً أنّ الشيفرة المصدريّة والملف الخاص بالطوبولوجيا الشبكيّة مُتاحَين على الرابط أَدناه:

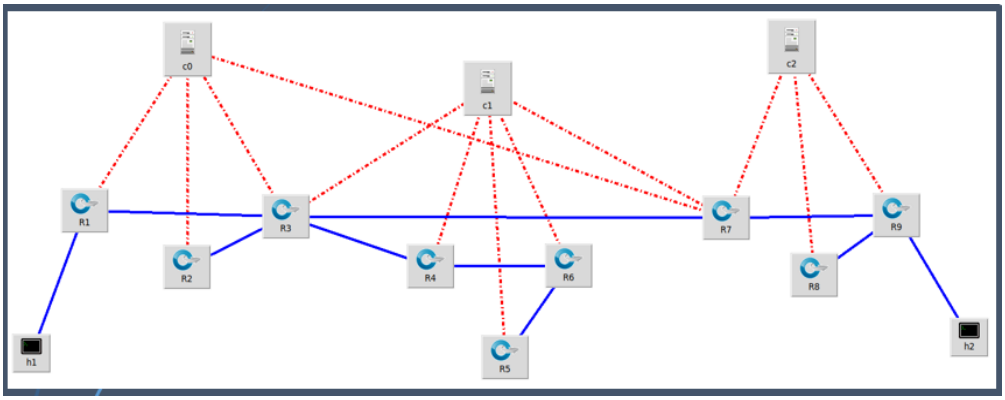
https://github.com/Ali-Mustafa96/Distributed_SD-NDN.git

٢,٦ بناء وتحقيق البُنية SD–NDN المُوزّعة

ستتضمّن البُنية المُوزّعة التي سنعتمدها ثلاث مُتحكِّمات (C0, C1, C2) وتسعة عُقد NDN Nodes (R1, R2, R3, … , R9) وعقدة المُستهلِك المُرسل للطلبات (Consumer 1) وعقدة المُنتج المُستقبِل للطلبات (Producer 1).

تم تحقيق هذه البُنية ضمن المُحاكي Mininet (بيئة IP–Based) وتشغيل الطوبولوجيا وإرسال بيانات بين عقدتي الإرسال والاستقبال كما هو مُحدّد ضِمن فقرة إعداد مُعلّمات المُحاكاة سابِقاً.

نُوضّح بالشكل (٥) هيئة الطوبولوجيا مُتعددة المُتحكِّمات (المُوزّعة) الناتجة داخل المُحاكي Mininet.



الشكل (٥) تحقيق البُنية SDN IP-Based التقليديّة المُوزّعة ضمن المُحاكي Mininet

ثمّ قُمنا بتحقيق هذه البُنية ضمن المُحاكي ndnSIM (أي ضمن البيئة NDN–Based) وتشغيل الشيفرة المصدريّة وعرض الطوبولوجيا الشبكيّة.

نُوضّح بالشكل (٦) الملف النصيّ الخاص بالطوبولوجيا والذي ستتم قراءتهُ في بداية الشيفرة المصدريّة الحاوية لإعدادات المُحاكاة. يحتوي هذا الملف ضِمن قِسمه الأول على جدول تعليمات لبِناء العُقد وتحديد إحداثياتها (x, y)، بينما يحتوي القسم الثاني على جدول تعليمات تعمل على إنشاء الوصلات بين العُقد وتحديد المواصفات الخاصة بها من سرعة وتأخيرات وحجم الأرتال.

```
# topo-tree-Distribute.txt

 * This scenario simulates a tree topology (using topology reader module)


router

#node   city    y   x   mpi-partition
leaf-1  NA  -90 -40 1
leaf-2  NA  -90 100 2
rtr-5   NA  -85 30  0
rtr-3   NA  -70 -10 1
rtr-4   NA  -70 20  3
rtr-6   NA  -70 40  2
rtr-8   NA  -70 70  4
rtr-1   NA  -60 -20 1
rtr-2   NA  -60 0   3
rtr-7   NA  -60 60  2
rtr-9   NA  -60 80  4
Cont1   NA  -45 30  0
Cont0   NA  -40 -10 1
Cont2   NA  -40 70  2

link

# from      to      capacity    metric  delay   queue
leaf-1      rtr-1   1Gbps       1   30ms    1000
rtr-3       rtr-2   1Gbps       1   30ms    1000
rtr-3       Cont0   1Gbps       1   30ms    1000
rtr-1       rtr-2   1Gbps       1   30ms    1000
rtr-1       Cont0   1Gbps       1   30ms    1000
rtr-2       Cont0   1Gbps       1   30ms    1000
Cont0       rtr-7   1Gbps       1   30ms    1000
rtr-2       Cont1   1Gbps       1   30ms    1000
rtr-2       rtr-7   1Gbps       1   30ms    1000
rtr-2       rtr-4   1Gbps       1   30ms    1000
Cont1       rtr-4   1Gbps       1   30ms    1000
Cont1       rtr-5   1Gbps       1   30ms    1000
Cont1       rtr-6   1Gbps       1   30ms    1000
Cont1       rtr-7   1Gbps       1   30ms    1000
rtr-4       rtr-6   1Gbps       1   30ms    1000
rtr-5       rtr-6   1Gbps       1   30ms    1000
rtr-4       rtr-6   1Gbps       1   30ms    1000
Cont2       rtr-7   1Gbps       1   30ms    1000
Cont2       rtr-8   1Gbps       1   30ms    1000
Cont2       rtr-9   1Gbps       1   30ms    1000
rtr-7       rtr-9   1Gbps       1   30ms    1000
rtr-8       rtr-9   1Gbps       1   30ms    1000
rtr-9       leaf-2  1Gbps       1   30ms    1000
```

الشكل (٦) بُنية الملف النصيّ الخاص بالطوبولوجيا المُوزّعة ضمن المُحاكي ndnSIM

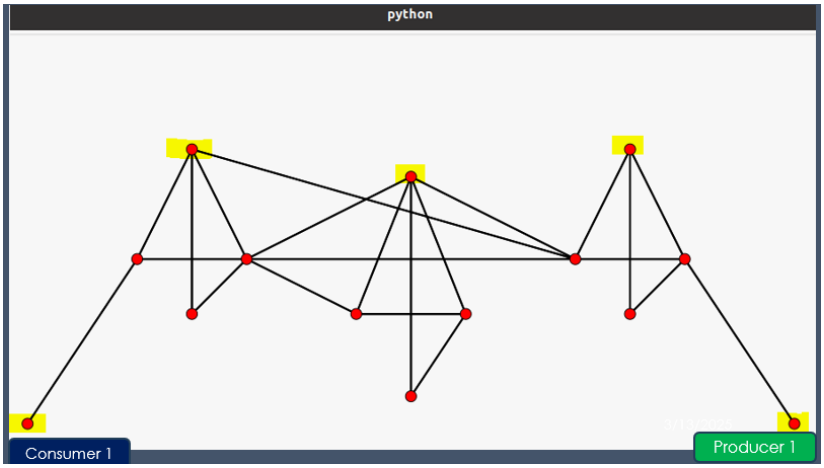
نعرُض بالشكل (٧) كيفيّة إضافة عُقد التحكُّم (لدينا ثلاث مُتحكّمات). حيث أنشأنا ثلاث أغراض من الموديول الفرعيّ "LearningController" (وهو نموذج برمجيّ يستخدم لبِناء المُتحكّمات) المُستورد من الموديول الأساسيّ "OpenFlow" ونربط هذه الأغراض مع العُقد التي قُمنا بقراءتها من ملف الطوبولوجيا مع تحديد ال Container المحتوية على العُقد الأُخرى (Consumer, Producer and NDN nodes).

```
135   Ptr<Node> switchNode1 = Names::Find<Node>("Cont0"); // The controller Cont0 node
136   Ptr<Node> switchNode2= Names::Find<Node>("Cont1");  // The controller Cont1 node
137   Ptr<Node> switchNode3= Names::Find<Node>("Cont2");  // The controller Cont2 node
138
139   // Create the switch netdevice, which will do the packet switching (Install OpenFlow )
140
141   OpenFlowSwitchHelper swtch1;
142   OpenFlowSwitchHelper swtch2;
143   OpenFlowSwitchHelper swtch3;
```

```
157     {
158       Ptr<ns3::ofi::LearningController> controller1 = CreateObject<ns3::ofi::LearningController> ();
159       if (!timeout.IsZero ()) controller1->SetAttribute ("ExpirationTime", TimeValue (timeout));
160       swtch1.Install (switchNode1, switchDevices, controller1);
161
162       Ptr<ns3::ofi::LearningController> controller2 = CreateObject<ns3::ofi::LearningController> ();
163       if (!timeout.IsZero ()) controller2->SetAttribute ("ExpirationTime", TimeValue (timeout));
164       swtch2.Install (switchNode2, switchDevices, controller2);
165
166       Ptr<ns3::ofi::LearningController> controller3 = CreateObject<ns3::ofi::LearningController> ();
167       if (!timeout.IsZero ()) controller3->SetAttribute ("ExpirationTime", TimeValue (timeout));
168       swtch2.Install (switchNode3, switchDevices, controller3);
169     }
170
```

الشكل (٧) إضافة عُقد المُتحكّمات ضِمن الشيفرة المصدريّة في المُحاكي ndnSIM

نُوضّح بالشكل (٨) شكل الطوبولوجيا المُوزّعة الناتجة ضمن المُحاكي ndnSIM بعد تشغيل الشيفرة المصدريّة الخاصة بها.



الشكل (٨) تحقيق البُنية SDN NDN-Based المُوزّعة ضمن المُحاكي ndnSIM

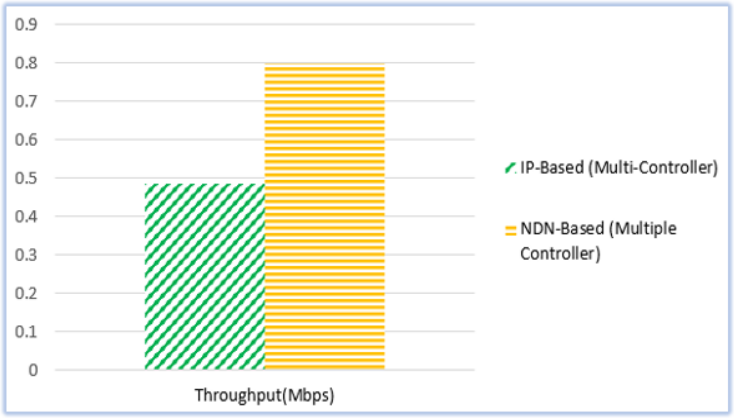
٧- اختبار أداء البُنية المُوزّعة ضمن البيئتين IP-Based و NDN-Based وتحديد التحسينات بعد تحقيق التكامُل مع بنية شبكات NDN في البُنية الجديدة:

سنقوم باختبار الأداء وِفق معايير الإنتاجية والتأخير وفقدان الرُزم واستهلاك عرض الحزمة. (قُمنا بتكرار عملية المُحاكاة عدة مرات وتمّ الحصول نتائج مُتقاربة)

تُعبّر الانتاجيّة عن كمية البيانات المُستلمة فعلياً بشكل سليم ضمن زمن مُحدّد، بينما يصِف التأخير الفترة الزمنيّة الكليّة حتى تتم معالجة وإيصال البيانات إلى وِجهتها، بينما فُقدان الرُزم (Packets Drop) تُعبر عن عدد الرُزم المُهملة (المفقودة) خلال الإرسال.

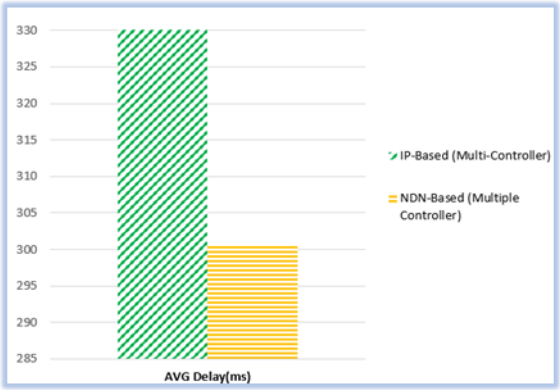
كما استخدمنا استهلاك عرض الحُزمة كمعيار أداء لإنّهُ يصِفُ مِقدار النطاق الترددي المطلوب أن تستهلكهُ البُنية الشبكية حتى تُنجز عملها. وبناءً على هذه المعايير سنُفاضِل بين البُنية SDN التقليديّة والبُنية SD-NDN.

قُمنا بإنجاز جميع التجارُب على المُحاكاة التي قُمنا بإعدادها وحصلنا على تحسُّن في الإنتاجية كما في الشكل (٩) بسبب قِلة الرُزم المفقودة التي لاحظناها في البيئة NDN– Based وسنوضّح ذلك عند عرضنا لمعيار فُقدان الرُزم.



الشكل (9) مُخطط يُوضّح الفرق بالإنتاجية ضمن البُنيتين IP-Based Multi-Controller

كما نعرِض ضمِن الشكل (١٠) التحسين الحاصِل بالتأخير (Delay) ضمن البيئة NDN-Based:



الشكل (١٠) مُخطط يُوضّح الفرق بالتأخير ضمن البُنيتين Multi-Controller

يعود التحسين من ناحية التأخير للأسباب التالية:

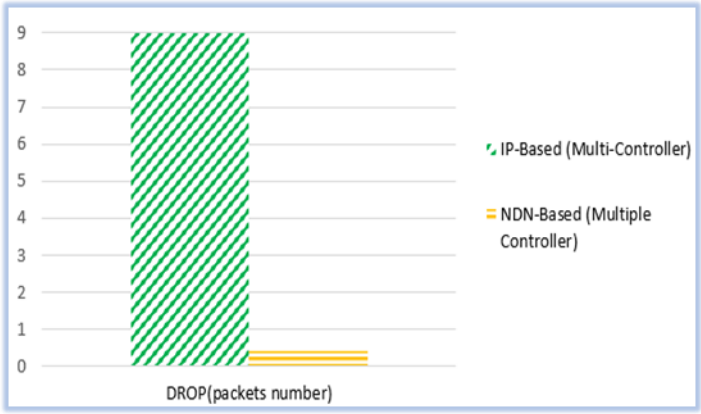
a. **السبب الأبرز هو التوجيه المُتكيّف (Adaptive Forwarding)**، حيث تدعم شبكة NDN إعادة التوجيه متعدد المسارات وفي حال حدوث أي خطب على وصلات الاتصال بين العُقد، يتم إعادة توجيه البيانات بشكل تلقائيّ عبر مسارات بديلة واختيار أفضل مسار ديناميكيًا دون انتظار وصول معلومات تحديث من مُتحكّم SDN بينما تعتمد شبكة SDN عادةً على مسارات مُعدّة مُسبقًا، مما يُسبب الزيادة في التأخير. **[10]**

b. **لا حاجة للاتصال نهاية إلى نهاية (End-to-End Connectivity)**، حيثُ تقوم شبكات NDN واعتماداً على أسماء المحتوى بعملية التوجيه بدلاً من العناوين المنطقيّة IP، مما يُلغي الحاجة إلى اتصالات نهاية إلى نهاية (End-to-End). يتم بذلك استرجاع البيانات من أي عُقدة تُخزّن المحتوى مؤقتًا، مما يُقلل من التأخيرات المرتبطة بإعدادات الاتصال. **[11]**

c. **لا يوجد ضِمن شبكات NDN أية أعباء إضافيّة مثل النموذج TCP/IP**، حيثُ يؤدي التخلّص من البروتوكول TCP/IP ضمن بيئة NDN إلى تقليل أزمنة

المُعالجة، بينما لا تزال شبكات SDN التقليديّة تستخدم بروتوكول TCP/IP، مما يتطلّب عمليات إضافية من أجل إعادة الإرسال ومعالجة الأخطاء والاختناقات والتدفّق. [12]

لاحظنا بالتجارب أيضاً قِلة فُقدان الرُزم ضمن البيئة NDN–Based كما في الشكل (١١) وقلنا أنّ هذا الأمر يُحسّن الإنتاجية.



الشكل (11) مُخطط يُوضّح الفرق بفقدان الرُزم ضمن البُنيتين Multi-Controller

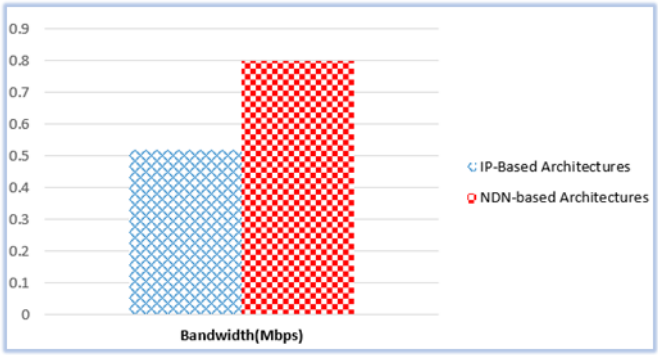
إنّ التحسُّن في إيصال الرُزم يعود لما يلي:

a) **التخزين المؤقت داخل الشبكة (In–Network Caching)**، حيثُ يُتيح التخزين المُؤقت داخل شبكات CCN استرجاع البيانات من العُقد الوسيطة، ويُقلّل من الحاجة إلى إعادة الإرسال من المصدر، الأمر الذي يُؤدي إلى تقليل فقدان الرُزم. [12]

b) **آلية الرُزم Interest/Data نفسها تُقلل من الازدحام وفقدان الرُزم**، حيثُ يتم ضمن شبكات NDN تبادل للبيانات فقط عند الطلب عن طريق رُزم ال Interest، مما يُقلل من نقل البيانات غير الضروريّ وتقليل فقدان الرُزم. بينما في شبكة SDN التقليديّة، ترسل البيانات عبر الوصلات إلى وجهتها، حتى لو لم يكُن بالإمكان

التعامل معها وتوجيهها فورًا، مما قد يؤدي إلى نفاذ مساحة التخزين المُؤقّت (Buffers) وارتفاع نسبة الفقد في الرُزم. **[11]**

على أية حال، إنّ البُنية SD–NDN تُعتبر أسوء من SDN المستند إلى IP (التقليديّة) من ناحية استهلاك النطاق الترددي (Bandwidth) ولاحظنا ذلك ضمن التجارُب ونتج لدينا الشكل (١٢) الذي يُوضّح الفرق بهذا الاستهلاك.



الشكل (١٢) مُخطط يُوضّح الفرق باستهلاك عرض الحزمة بين البيئتين IP-Based و NDN-Based

ترتفع نسبة استهلاك عرض الحُزمة ضمن بُيئة SDN المُعتمدة على NDN بسبب:

A. تستهلك حركة مرور الخاصة بعمليات مزامنة ذاكرة التخزين المُؤقت (Cache Synchronization Traffic) عرض حُومة أكبر، حيثُ تقوم شبكات NDN بآليات مُزامنة ذاكرة التخزين المؤقت (للحِفاظ على الاتّساق بالمُحتوى بين العُقد NDN) الأمر الذي يدعو لوجود بيانات تحكُّم إضافية. تستهلك تحديثات ذاكرة التخزين المؤقت في شبكةNDN عرض حُزمو إضافيّ، خاصةً في الشبكات الديناميكية والكبيرة. **[13]**

B. تزيد بادئات الأسماء الكبيرة من حجم الرُزمة (Large Name Prefixes)، حيثُ تستخدم NDN أسماء المحتوى بدلاً من عناوين IP، وبالتالي مُمكن أن تحتوي ترويسات الرُزم على مسارات تسمية هرميّة طويلة. بينما نعلم أنّ آلية العنونة تكون ذات حجم ثابت في الأنظمة IP–Based. **[13]**

تمت تقديم التوصية ضمن شبكات البُنية الجديدة SDN–Based Internet2 (٢٠١٤+) مع شبكات NDN بأن يكون عرض الحزمة كافي ومن مرتبة كبيرة تصِل إلى ١ Gbps وما فوق، وبالتالي فإنّ استهلاك عرض الحزمة لن يكون تحدٍ كبير عند تحقيق التكامُل بين البُنيتين.

٨- الخاتِمة والأعمال المُستقبليّة

قدّمنا في هذا البحث نموذج وتحقيق فعليّ للبُنية الجديدة SD-NDN مُتعدّدة المُتحكّمات وقُمنا بتحديد ما يُناسِبُها من مُعلّمات مُحاكاة، وقد ضمَّ هذا التحقيق كِلا البيئتين IP-Based التقليديّة و NDN-Based الجديدة، كما قُمنا بتحديد كل التحسينات التي سبّبتها البُنية NDN ضمن هذه البُنية الجديدة وعمِلنا على قياس أهم معايير الأداء ووضع التفسيرات المُناسبة لكل تحسّن تم إيجادُه.

نتطلّع مُستقبلاً إلى التطوير على الشِيفرة المصدريّة الخاصة بالبُنية SD-NDN المُوزّعة بحيث تُصبح داعِمة لتخصيص خوارزميات مُوازنة الحِمل بين المُتحكُّمات وتحديد أي وسطاء تتعلّق بذلك. يمكن أن تُجرى أعمالاً على هذا البحث مثل تطبيق البُينة-SD NDNعلى بُنية مُوزّعة تشابكية هرميّة، وتطوير الشيفرة المصدريّة للبُنية المُوزّعة حتى تدعم (بالإضافة إلى خوارزميات مُوازنة الحِمل) تحديد نسبة تدفّقات الرُزم ضِمن كل مُتحكُّم.

٩- جدول المُختصرات:

CCN	Content-Centric Network
CCP	Cache replacement policy based-on Content Popularity
CS	Content Store
FIB	Forward Information Base
ICN	Information-Centric Network
NCP	Network Control Protocol
NDN	Named Data Network
PIT	Pending Interest Table
SDN	Software-Defined Network
SD-NDN	Software-Defined Named Data Network

١٠- المراجِع:

[١] A. Marica ،C. Claudia ،R. Giuseppe ،M. Antonella و I. Antonio ، "Understanding Name-based Forwarding Rules in Software-Defined Named Data Networking "،*University of New South Wales. Downloaded on September 26,2020 at 18:20:29 UTC from IEEE Xplore* .٢٠٢٠ ،.

[٢] L. R. M. André ،R. C. Francisco ،B. A. Guilherme و N. S. Leobino ، "NDN Fabric: Where the Software-Defined Networking Meets the Content-Centric Model "،*IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, VOL. 18, NO. 1* ،March 2021 .

[٣] Y. Sembati ،N. Naja و A. Jamali ،"A review of SDN-enabled routing protocols for Named Data Networking "،*Engineering Reports published by John Wiley & Sons Ltd* ٦ ،.March 2024 .

[٤] A. Albatool و B. Abdelfettah ،"Software-Defined Named Data Networking in Literature: A Review "،*Future Internet 2024, 16, 258* ،. ٢٣July 2024 .

[٥] C. Lorenzo ،H. Mariann ،L. Lars ،S. Niranjan و T. Mauro ، "Considerations on the Adoption of Named Data Networking (NDN) in Tactical Environments "،*International Conference on Military Communications and Information Systems (ICMCIS)* ، ٢٠١٩.

[٦] M. Marion ،Development of a performance measurement tool for SDN ،٢٠١٥.

[٧] S. S. M.Micheal و Dr.K.Kumar. ،"A SURVEY ON NAMED DATA NETWORKING "،*IEEE SPONSORED 2ND INTERNATIONAL CONFERENCE ON ELETRONICS AND COMMUNICATION SYSTEM* ، ٢٠١٥.

[٨] J. Zain ul Abidin ،A. Zeeshan و T. Muhammad ،"A Named Data Networking (NDN), New Approach to Future Internet Architecture Design: A Survey "،*International Journal of Informatics & Communication Technology (IJ-ICT) Vol.2, No.3* ،Desemper 2013 .

[٩] د. ي. غازي و م. ع. الحمصيه، "سياسة مُحسّنة لاستبدال ذاكرة التخزين المُؤقت في شبكات البيانات المسمَّاة،" *مجلة جامعة حمص المجلد ٤٦ العدد ٥*، ٢٠٢٤.

[١٠] S. Kalafatidis ،S. Sotiris ،D. Vassilis ،M. Lefteris و T. Vassilis ، "Logically-Centralized SDN-Based NDN Strategies for Wireless Mesh Smart-City Networks "،*Future Internet 2023, 15, 19. https://doi.org/10.3390/fi15010019* ٢٩ ،December 2022 .

[١١] "Named Data Networking،" Wikipedia article.

[١٢] R. S. Nana ،T. Harashta ،M. Alvin و P. Fajar ،"Performance Comparison of Named Data and IP-based Network—Case Study on the Indonesia Higher Education Network "،*Journal of Communications Vol. 13, No. 10* ،October 2018 .

[١٣] A. Manar ،A.-A. Dawood ،A. Medhat و K. Firdous ،"Leveraging ICN and SDN for Future Internet Architecture: A Survey "،*Electronics 2023, 12, 1723. https://doi.org/10.3390/electronics12071723* ٤ ، April 2023 .

[١٤] X. Wenfeng ،W. Yonggang ،S. M. IEEE ،H. F. Chuan ،S. M. IEEE ،N. Dusit ،M. IEEE ،X. Haiyong و M. IEEE. ،"A Survey on Software-

Defined Networking "،*.COMMUNICATION SURVEYS & TUTORIALS, VOL. 17, NO. 1* ،FIRST QUARTER 2015 .

[١٥] و POXم. الضاهر، "تحليل أداء وحدات تحكم الشبكات المعرفة برمجياً: Opendaylight .٢٠٢١ ،*مجلة جامعة حمص المجلد ٤٣ العدد ١٧*، "،

[١٦] "https://ndnsim.net/current/intro.html .[متصل] "،

تطوير PopScore كمقياس يعتمد على تحليل التشابه لتقييم نجاح المنشورات التسويقية على منصة الفيسبوك

المهندسة راما أكرم ياسين
(قسم هندسة البرمجيات، كلية الهندسة المعلوماتية، جامعة حمص، حمص، سورية - ٢٠٢٤)
إشراف: الدكتور ناصر أبو صالح – الدكتور وسيم رمضان

الملخص

تُعتبر منصات التواصل الاجتماعي من أهم أدوات التسويق الإلكتروني في العصر الحالي، حيث تُتيح للمسوقين الوصول إلى جمهور واسع والتفاعل معه بطرق مبتكرة وفعالة. على الرغم من ذلك، لا توجد دراسات كافية تُحدد العوامل المؤثرة على فعالية الاستراتيجيات التسويقية المتبعة وتأثيرها على سلوك المستخدم في منصات التواصل الاجتماعي.

تهدف هذه الدراسة إلى تطوير مقياس جديد PopScore للتعبير عن درجة انتشار وشعبية المنشور الجديد المتوقعة قبل إطلاقه على منصة فيسبوك، مما يساهم في تقديم رؤية مسبقة للمسوقين حول فعالية المنشور ونجاحه وذلك بالاعتماد على دمج عدة تقنيات من تحليل البيانات ومعالجة النصوص من أجل دراسة وتحليل المنشورات المشابهة التي تم نشرها سابقاً وكيفية تفاعل المستخدمين معها من خلال تطبيق مفهوم الفلترة بحسب المحتوى Content Filtering ومن ثم إيجاد أفضل معادلة لحساب الانتشار والتي تعطي نتائج أدق في عملية التنبؤ، وبالتالي القدرة على تقييم فعالية استراتيجيات التسويق الإلكتروني المتبعة على منصات التواصل الاجتماعي.

تم التركيز على منصة فيسبوك، نظراً لانتشارها الواسع واستخدامها من قبل الشركات والمسوقين للترويج لمنتجاتهم وخدماتهم، وساهمت هذه الدراسة في سد الفجوة في القدرة على

معرفة مدى نجاح المنشور التسويقي قبل إطلاقه وبالتالي فهم كيفية تحقيق أقصى استفادة من منصات التواصل الاجتماعي في عمليات التسويق الإلكتروني والحملات الإعلانية، حيث بلغت دقة المقياس في التنبؤ 0.977.

الكلمات المفتاحية: الفلترة حسب المحتوى، تشابه، تسويق إلكتروني، تواصل اجتماعي، فيسبوك، تقييس، تصنيف.

Formulating PopScore: A Similarity-Based Metric for Evaluating the Success of Marketing Posts on Facebook

Eng. Rama Akram Yaseen
(Department of Software Engineering, Faculty of IT, Al-Baath University, Homs, Syria - 2024)
Dr. Naser Abo Saleh – Dr. Wassim Ramadan

Abstract

Social media platforms are considered one of the most important digital marketing tools in the current era, as they allow marketers to reach a wide audience and interact with them in innovative and effective ways. Despite this, there are not sufficient studies that determine the factors affecting the effectiveness of the marketing strategies used and their impact on user behavior on social media platforms.

This study aims to develop a new metric, PopScore, to express the expected score of spread and popularity of a new post before it is published on Facebook. This metric provides marketers with an early insight into the effectiveness and success of the post by integrating several data analysis and text processing techniques to study and analyze similar previously published posts and user interactions with them. This approach applies the concept of Content Filtering to find the best formula for calculating reach, which yields more accurate

prediction results, thereby enabling the evaluation of the effectiveness of digital marketing strategies on social media platforms.

The focus was on Facebook due to its widespread use by companies and marketers to promote their products and services. This study fills the gap in understanding the success potential of a marketing post before its launch, thus understand how to make the most of social media platforms in electronic marketing operations and advertising campaigns. The model's accuracy rate in prediction reached 0.977.

Keywords: Content Filtering, Similarity, Digital Marketing, Social Media, Facebook, Normalization, Classification.

١. المقدمة

تُعد منصات التواصل الاجتماعي من أكثر الظواهر التكنولوجية انتشاراً في القرن الحادي والعشرين، حيث أنها أصبحت جزءاً لا يتجزأ من حياة الملايين وأتاحت للأفراد إمكانية التواصل الفوري وتبادل الأفكار والمحتوى بمختلف أنواعه بسهولة غير مسبوقة.

تميزت هذه المنصات بالقدرة على خلق شبكات اجتماعية رقمية تربط بين الأشخاص من مختلف بقاع الأرض، ومن ثم توسعت لتشمل مختلف جوانب الحياة مثل العمل، التسوق، التعلم، والترفيه.

تبين من خلال دراسة أُجريت عام 2013 [1] أن 81% من المسوقين يستخدمون وسائل التواصل الاجتماعي كأهم أداة تسويقية لهم.

يُعزى الانتشار السريع لوسائل التواصل الاجتماعي إلى عدة عوامل، من أبرزها التطور التكنولوجي السريع في الهواتف الذكية وتوافر الانترنت بأسعار معقولة، مما جعل الوصول إليها سهلاً ومتاحاً للجميع، حيث بلغ عدد مستخدمي الانترنت بحسب دراسة أجريت عام 2022 حوالي ثلاثة مليارات مستخدم حول العالم.[2]

ومع انتشار الأوبئة المعدية في الآونة الأخيرة مثل كورونا وغيرها أدى ذلك إلى عرقلة حركة الأسواق وتراجعها ونتيجة لذلك توجه العديد من التجار إلى التسويق عبر الانترنت [3] مستغلين بذلك التواجد الكبير للأشخاص على منصات التواصل الاجتماعي وتم اتخاذها كأداة أساسية في عالم التسويق الإلكتروني، حيث أتاحت للتجار الوصول إلى جمهور واسع يتعدى الحدود المكانية والجغرافية والتفاعل معه بطرق مبتكرة وفعالة من خلال التواصل المباشر مع الزبائن، وتحليل البيانات لتحسين استراتيجياتها التسويقية.

تختلف نسبة المبيعات بين كل مسوّق وآخر بسبب اختلاف أساليب عملية طرح المنتج والترويج له. وبالتالي نحن نشهد محاولات وأساليب تسويقية كثيرة لكن لم يتم إثبات أيٍّ منها كان الأجدر والأكثر رواجاً حتى الآن. حيث أن الأبحاث ركزت في المقام الأول على تأثير الجائحة على سلوكيات الصحة الوقائية ولا يوجد إلا القليل من الاهتمام لتأثير الوباء على السلوكيات الشرائية في وسائل التواصل الاجتماعي [3].

ومع أن التسويق الإلكتروني شهد اهتماماً بحثياً كبيراً، إلا أن الدراسات السابقة ركزت في الغالب على مواقع التجارة الإلكترونية، مما يجعل نتائجها غير قابلة للتعميم على منصات التواصل الاجتماعي بسبب اختلاف طبيعتها. وهنا ظهرت لدينا فجوة حول التحديات المتعلقة بتقييم فعالية المنشورات التسويقية على منصات التواصل الاجتماعي.

لمواجهة هذه الفجوة، تم تطوير عدة حلول بدءاً من الأساليب التقليدية القائمة على خبرة الخبراء وحتى استخدام تقنيات التعلم الآلي الحديثة. في البداية، اعتمد الخبراء على التحليل اليدوي لمقاييس الأداء التقليدية مثل معدل النقر إلى الظهور (CTR) وهو نسبة الأشخاص الذين نقروا على الرابط أو الإعلان مقارنةً بالأشخاص الذين شاهدوا الإعلان، ومعدل التحويل (Conversion Rate) وهو نسبة الأشخاص الذين قاموا بالإجراء مثل الشراء أو التسجيل بعد النقر على الإعلان لتحديد مدى فعالية الإعلان، بالإضافة إلى ذلك، تم استخدام استطلاعات الرأي واختبارات A/B لجمع بيانات مباشرة من المستهلكين حول فعالية المنشورات المختلفة.

على الرغم من أهمية هذه الأساليب، إلا أنها قد تكون محدودة في قدرتها على التنبؤ بالأداء المستقبلي بدقة عالية. ومع التقدم التكنولوجي وتوفر كميات ضخمة من البيانات، أصبح من الممكن تطبيق تقنيات تحليل البيانات ومعالجة النصوص بهدف تقييم فعالية المنشورات التسويقية بشكل أكثر دقة. مثلاً يمكن أن يتم استخدام تقنية TF-IDF لاستخراج الميزات النصية من المنشورات مما يساعد في التنبؤ بمدى نجاح المنشور قبل نشره. كذلك، يمكن أن تُستخدم تقنيات التجميع مثل K-Means لتقسيم المنشورات إلى مجموعات بناءً على خصائصها المشتركة، مما يسهل تحليل أنماط التفاعل وتحسين الاستراتيجيات التسويقية.

علاوةً على ذلك، تُعد نماذج التنبؤ مثل الانحدار الخطي والشبكات العصبية العميقة أدوات قوية لتقدير نجاح المنشورات المستقبلية. هذه النماذج تعتمد على بيانات تاريخية لتحليل الأنماط والتنبؤ بالأداء المستقبلي، مما يتيح للتجار والشركات اتخاذ قرارات مبنية على بيانات دقيقة بدلاً من الاعتماد على الحدس والخبرة الشخصية فقط.

تعتبر هذه الحلول المتقدمة جزءاً من التطور المستمر في مجال التسويق الإلكتروني، حيث تسهم في تقديم رؤى أعمق حول سلوك المستخدمين وتحسين فعالية الاستراتيجيات التسويقية على منصات التواصل الاجتماعي.

في هذه الدراسة، سيتم التركيز على تطوير مقياس جديد يمكنه تقدير درجة انتشار المنشورات التسويقية قبل نشرها، اعتماداً على تحليل البيانات ومعالجة النصوص لفهم العوامل المؤثرة على تفاعل المستخدمين. ويمثل هذا المقياس إضافة نوعية للتقنيات الحالية من خلال توظيف أدوات تحليل التشابه بين المنشورات واستنباط الأنماط الفعّالة في التفاعل معها، مما يدعم تحقيق الاستراتيجيات التسويقية الأمثل على منصات التواصل الاجتماعي.

تم استخدام خوارزمية TF-IDF لاستخراج الميزات النصية من المنشورات، بالإضافة إلى تحليل تشابه جيب التمام Cosine Similarity لقياس مدى التشابه بين المنشورات المختلفة. يتم تقييم المنشورات الجديدة بناءً على تشابهها مع منشورات سابقة من حيث تفاعل المستخدمين معها، ويتم تصنيفها ضمن فئات مختلفة تعبر عن شعبية المنشور وفقاً لقيمها التنبؤية.

من أجل التقييم تم استخدام مقاييس التنبؤ مثل متوسط الخطأ المطلق (MAE)، متوسط مربع الخطأ (MSE)، جذر متوسط مربع الخطأ (RMSE) لتقييم دقة درجة الانتشار التنبؤية PopScore بالإضافة إلى مقياس F1 لتصنيف شعبية الانتشار Popularity. وقد ساعدت هذه المقاييس في تحديد مدى قدرة المنهجية المتبعة في التنبؤ بنجاح المنشورات قبل إطلاقها وحققت نتائج واعدة من حيث دقة التنبؤ بدرجة الانتشار وتحديد الشعبية.

٢. مشكلة البحث

على الرغم من وجود العديد من الدراسات التي تناولت مواضيع التسويق الإلكتروني إلّا أنَّ معظمها كان قائماً على مواقع التجارة الإلكترونية ولا يمكن إسقاط كافة نتائجها على منصات التواصل الاجتماعي نتيجة لاختلاف بيئة وطبيعة تدفق المعلومات بين مواقع التجارة الإلكترونية ومنصات التواصل الاجتماعي. لذلك لايزال لدينا قصور في دراسة هذا المجال من خلال منصات التواصل الاجتماعي ولا سيما منصة فيسبوك نتيجة سياستها الصارمة

في الحصول على بيانات قابلة للتحليل والدراسة. وهذا القصور في البحث يترك فجوة في الفهم العميق لكيفية تحقيق أقصى استفادة من هذه المنصات، وتأثيراتها المختلفة على سلوك المستهلك، وفعالية الاستراتيجيات المتبعة.

٣. هدف البحث

يهدف البحث إلى إيجاد طريقة ذكية لقياس فعالية المنشورات التسويقية على وسائل التواصل الاجتماعي قبل إطلاقها، وذلك باستخدام تقنيات تحليل البيانات ومعالجة النصوص المتطورة. من خلال استخلاص مقياس جديد يسمى PopScore لقياس درجة انتشار المنشور المتوقعة من أجل تحقيق تسويق إلكتروني فعال.

في إطار هذه الدراسة، سيتم تحليل أنواع مختلفة من المنشورات على منصة فيسبوك، بما في ذلك منشورات الصور، مقاطع الفيديو والمنشورات النصية. سيتم تقييم سلوك المستخدمين من خلال متابعة عدد المشاركات، التفاعلات، والتعليقات على كل نوع من هذه المنشورات، بهدف تحديد العوامل المؤثرة على تفاعل المستخدمين وتقييم فعالية المنشورات.

٤. أهمية البحث

تكتسب هذه الدراسة أهميتها من مساهمتها في سد فجوة البحث الحالية في مجال التسويق الإلكتروني على منصات التواصل الاجتماعي، حيث توفر للمسوقين أداة تحليلية تُمكنهم من التنبؤ بفعالية منشوراتهم قبل نشرها. كما أن المنهجية المقترحة تعتمد على أدوات تحليلية متقدمة تساعد في فهم أعمق لكيفية تفاعل المستخدمين مع المنشورات. بذلك، تتيح الدراسة اتخاذ قرارات تسويقية تعتمد على بيانات دقيقة، مما يعزز القدرة على تحسين استراتيجيات التسويق وتحقيق نتائج أفضل في الحملات الإعلانية.

٥. الدراسات المرجعية

تكشف أدبيات الدراسات التي تمت مراجعتها عن مجموعة متنوعة من الاستراتيجيات والتقنيات المستخدمة لتقييم كيفية تحقيق تسويق إلكتروني فعّال على منصات التواصل الاجتماعي. بدءاً من تحليل التغريدات ودراسة استخدام الهاشتاغات الرائجة على تويتر لتحليل مشاعر الجمهور واستجابته لموضوع معين [4] ، حيث أظهر التجميع الموضوعي للهاشتاغات دقة أعلى في تحديد مشاعر وتجاوب المتابعين تجاه موضوع معين سواء تسويقي أو اجتماعي أو رأي عام مقارنة بالتحليل غير المجمّع.

إلى استخدام تقنية التجميع K–means في دراسة سلوك المستهلك لشركة كوكاكولا عبر الفيسبوك [5] ، التي أبرزت أهمية معدل استخدام المنتج كعامل مؤثر في سلوك المستهلك وتفاعله.

كما تناولت دراسة أخرى تحليل وجهات نظر المستهلكين حول استخدام بيانات وسائل التواصل الاجتماعي لأغراض تسويقية [6] ، وأظهرت أن الإعلانات المستهدفة تُعد الأكثر فعالية، حيث يمكن أن تزيد المبيعات بأكثر من ١٠٪ على الرغم من أن غالبية المستجيبين لم يكونوا مرتاحين لاستخدام المسوقين لبياناتهم العامة.

وأكدت دراسة تحليل تغيرات مكانة العلامة التجارية وتفاعل الزبائن معها [7] أن نوعية المنشورات ومحتواها هي العامل الأساسي في زيادة التفاعل وليس عدد المنشورات الكبير.

إضافة إلى ذلك، ركزت دراسة توقع سلوك تفاعل العملاء مع المنشورات التسويقية باستخدام تقنيات التعلم الآلي [8] على أهمية خصائص المنشورات مثل صياغته وطوله وتوقيت نشره والتساؤلات الموجودة فيه في زيادة التفاعل، في هذه الدراسة أظهرت الشبكات العصبية أفضل أداء في التنبؤ بسلوك التفاعل من حيث الإعجاب أو التعليق أو المشاركة.

كما أكدت دراسة تأثير البث المباشر على اتخاذ قرارات المستهلك [9] إلى أن تفاعل المستخدمين الآخرين ونشرهم لفيديوهات تظهر فك التغليف أو فيديوهات لتقييم المنتج بمرحلة ما بعد الشراء له تأثير كبير على قرار المستهلك لأنه يُظهر التجربة الواقعية والمراجعات الموثوقة وبالتالي يؤثر بشكل إيجابي على نوايا الشراء.

بالإضافة إلى ما سبق، قدمت دراسة مسحية شاملة [10] تحليلاً لمجموعة من المقالات التي تناولت تقنيات التصنيف على وسائل التواصل الاجتماعي، حيث أظهرت أن تويتر هو المنصة الأسهل لاستخراج البيانات وأن تقنية K–means هي الأبسط والأكثر شيوعاً. كذلك، أظهرت الدراسة فعالية تحليل المشاعر عند دمجه مع تقنيات التصنيف في استخراج رؤى تسويقية قيمة.

أما الدراسة الأخيرة [11] فقد ركزت على وضع رؤى استراتيجية ومبادئ توجيهية لزيادة فعالية وسائل التواصل الاجتماعي. وتناولت تأثير العوامل المتعلقة بالمنشورات، والعوامل السياقية والزمنية، واختلافات الصناعات والمنصات على تفاعل المستخدمين. وأظهرت الدراسة أن الأنواع المختلفة لمحتوى المنشور والتي تحدد فيما إذا كان إعلامي أو ترفيهي أو إعلاني ترويجي ونوع المنصة يلعبان دوراً كبيراً في تقديم الرؤى الاستراتيجية وأنه لا يمكن تعميم الإرشادات على كافة العلامات التجارية بشكل دائم.

يوجد العديد من الدراسات التي تناولت موضوع التسويق الإلكتروني ولكن معظمها كان قائماً على مواقع التجارة الإلكترونية وأن القليل منها تطرق إلى التسويق عبر منصات التواصل الاجتماعي حيث أن التطبيقات الأكثر شيوعاً كانت عبارة عن تحليل المشاعر، والتنقيب عن الرأي واكتشاف التوجه، على الرغم من أنه حالياً يتم التوجه في البحث العلمي للتسويق عبر هذه المنصات والواضح أنه يتم إعطاءها اهتمام من قبل الباحثين ولكن ما يزال هناك قصوراً وحاجة إلى المزيد من الدراسات، كما أن معظم الدراسات كانت تركز على اللغة الإنكليزية والسوق الأجنبي بشكل عام ولا يوجد تطرق كبير لدراسة السوق العربي عامةً والسوري خاصةً، بالإضافة إلى التركيز على منصة تويتر ويعود ذلك إلى سهولة الحصول على البيانات منها مقابل سياسة فيسبوك الصارمة بذلك.

بيَّنت الدراسات السابقة أهمية وسائل التواصل الاجتماعي في التسويق الإلكتروني، إلا أنها ركزت بشكل أكبر على تحليل المشاعر وتحديد الجمهور المستهدف. في المقابل، تسعى هذه الدراسة إلى سد هذه الفجوة من خلال تطوير طريقة جديدة لقياس فعالية المنشورات التسويقية قبل نشرها، وذلك بدمج تقنيات تحليل البيانات ومعالجة النصوص. وبالتالي، فإن هذه الدراسة تساهم في تعميق فهمنا للعوامل التي تؤثر على نجاح الحملات التسويقية في منصات التواصل الاجتماعي.

٦. أدوات وطرائق البحث

٦,١ مجموعة البيانات

تحتوي مجموعة البيانات المستخدمة في هذا البحث على منشورات من منصة التواصل الاجتماعي – فيسبوك – حيث أن هذه المنشورات تغطي الفترة الزمنية الممتدة من منتصف 2017 وحتى آب 2023.

مصدر البيانات:

تعود هذه البيانات إلى شركة الكرم للتجهيزات السنية، وهي شركة سورية مقرها الرئيسي في حلب لكنها تملك عدة أفرع في المحافظات السورية وهي وكيل لعدة علامات تجارية عالمية، تم جمع المنشورات باستخدام واجهة برمجة التطبيقات API الخاصة بمنصة Facebook،

واستهلكت عملية الجمع الكثير من الوقت والجهد نتيجة حظر الوصول إلى واجهة برمجة التطبيقات من سوريا.

تم الحصول على بيانات المنشورات من فيسبوك وفقاً للتراخيص والشروط الخاصة بالمنصة، بعد أن تم الحصول على موافقة شركة الكرم للتجهيزات السنية بالوصول إلى بيانات الصفحة بشرط أن يتم استخدامها لتخديم هذا البحث فقط وعدم استخدامها أي استخدامات أخرى خارج هذا البحث.

وصف البيانات:

تتميز المنشورات في مجموعة البيانات بتنوعها الكبير، حيث تشمل مختلف أشكال المحتوى، مثل:

- **النص:** يتمثل النص في محتوى مكتوب يعبر عن أفكار أو معلومات معينة.
- **الصور:** تشمل الصور رسومات ثابتة تعبر عن أفكار أو مشاعر أو معلومات بطريقة بصرية.
- **الألبومات:** تتكون الألبومات من مجموعة من الصور المرتبطة ببعضها البعض، وتروي قصة أو تقدم عرضاً متسلسلاً.
- **الفيديوهات:** تشمل الفيديوهات تسجيلات متحركة تعبر عن أفكار أو معلومات أو مشاعر بطريقة ديناميكية وجذابة.
- **البث المباشر:** يتيح البث المباشر للمستخدمين مشاركة لحظاتهم مع متابعيهم بشكل مباشر، مما يخلق تفاعلاً تلقائياً.
- **ريل:** يُعد الريل نوعاً حديثاً من الفيديوهات القصيرة ذات المحتوى الترفيهي أو التثقيفي أو التسويقي.

الميزات الخام التي قمنا بالحصول عليها من API المنصة كان البعض منها عبارة عن أغراض json متداخلة تحتاج إلى معالجة لاستخراج القيم منها وبعضها الآخر عبارة عن قيم مباشرة، بالاعتماد على هذه الميزات تم استخراج واشتقاق العديد من الميزات الأخرى من خلال عمليات التحليل الإحصائي ومعالجة النصوص ليصبح لدينا عدد أكبر من

الميزات التي يمكن دراستها في عملية تحليل المنشور من حيث المحتوى والموضوع والتفاعل.

تم إدراج ميزات المنشور النهائية المعالجة في ***الجدول رقم (1)***. حيث كل منشور يمثل سجلاً في مجموعة البيانات.

الجدول رقم (1): توصيف لأعمدة المنشور الموجودة في مجموعة البيانات.

اسم العمود	نوعه	أساسي أم مشتق؟	محتواه	قيمه الفريدة
id	object | string	أساسي	يمثل الرقم التعريفي للمنشور وهو قيمة فريدة	
created_time	object | datetime	أساسي	يمثل زمن نشر المنشور	
permalink_url	object | string	أساسي	يمثل رابط المنشور	
shares	Int64	أساسي (على شكل Json)	عدد مرات مشاركة المنشور	
comments_count	int64	أساسي	عدد التعليقات الأساسية على المنشور بدون الردود	
reactions_count	int64	أساسي	عدد التفاعلات على المنشور	
message	object | string	أساسي	يمثل المحتوى النصي للمنشور بالكامل مع الهاشتاغات	
message_without_#	object | string	مشتق	يمثل المحتوى النصي للمنشور بدون الهاشتاغات	
message_without_#_length	int64	مشتق	طول المحتوى النصي للمنشور بدون الهاشتاغات	

True | False	هل message_without #_ يحتوي على محارف عربية أم لا	مشتق	bool	msg_is_ar
	طول المحارف العربية	مشتق	int64	msg_ar_len
True | False	هل message_without #_ يحتوي على محارف انكليزية أم لا	مشتق	bool	msg_is_en
	طول المحارف الانكليزية	مشتق	int64	msg_en_len
'cover_photo' 'profile_media' 'photo' 'album' 'share' 'video_inline' 'video_direct_respons' 'text' 'native_templates' 'video_autoplay' 'event' 'video' 'scheduled_live_video_post 'messenger_generic_templa '	نوع المنشور	مشتق من attachments وهو عبارة عن غرض json يحوي جميع الوسائط الموجودة في المنشور من صور وفيديوهات وألبومات.	object	type
	عدد عناصر الميديا في حالة أن المنشور يحوي ألبوم	مشتق من attachments	int64	subattachments_count
True | False	تمثل احتواء المنشور على الكلمات الأساسية أم لا	مشتق	bool	contains_main_topics

distinct_emojis		مشتق	تمثل قائمة الرموز التعبيرية الموجودة في المنشور	['🤩', '📣']
emoji_count	int64	مشتق	عدد الرموز التعبيرية	
unique_emoji_count	int64	مشتق	عدد الرموز التعبيرية الفريدة	

ضمَّت مجموعة البيانات الخام 1819 منشوراً، تم حذف المنشورات المكررة بالاعتماد على الرقم التعريفي للمنشور ليصبح لدينا 1641 منشوراً، ومن ثم وجدنا أنه لدينا عدة منشورات عائدة من الAPI لا تحوي على أية قيمة في الـ message ولا في الـattachments وبالتالي فإن محتواها غير معروف لذلك تم حذفها ليصبح العدد النهائي 1632 منشوراً. كما أن ميزة عدد المشاركات كانت على شكل غرض json تحوي الخاصية { count:numeric_value }، بعض المنشورات كانت ميزة عدد المشاركات لها فارغة، تم زيارة روابط هذه المنشورات يدوياً لنجد أن جميعها كانت لا تحوي أي مشاركات، لذلك تم تعبئة القيم الفارغة لهذه الميزة بقيمة صفر.

عينة من البيانات النهائية المستخدمة بعد المعالجة:

id: ١٨٤٧٨٠٧٠٨٨٨٨٠٠٧٠_3410072262653537

created_time: 2022-11-29T14:54:07+0000

permalink_url: https://www.facebook.com/1847807088880070/posts/3410072262653537/

shares: 14

comments_count: 399

reactions_count: 207

message:
لأن رضاكم وثقتكم دائماً من أولوياتنا ..
عملنالكم العرض #عرضين
عرض من #شركة_الكرم وعرض الـ Black Friday من #Eighteeth
العرض ساري لغاية يوم الجمعة ٢٠٢٢/١٢/٢
تفاصيل Fast-fill:
/alkaramdent.com/product/fast-fill
تفاصيل Fast-pack:
/alkaramdent.com/product/fast-pack

message_without_#:	تفاصيل Ultra X: https://alkaramdent.com/product/ultra-x/ انستغرام: https://instagram.com/alkaramdentalsupplies #YouDeserveTheBest #AlkaramDentalSupplies #Eighteeth #Devices #materials #dental #treatment #dentistry #clinic #offers #أجهزة_سنية #مواد_سنية #مؤتمر #بلاك_فرايدي #عروض #أطباء_الأسنان Black عرض من وعرض الـ لأن رضاكم وثقتكم دائماً من أولوياتنا .. عملنالكم العرض Fast-fill: من العرض ساري لغاية يوم الجمعة ٢٠٢٢/١٢/٢ تفاصيل Friday Fast-pack: تفاصيل alkaramdent.com/product/fast-fill/ Ultra X: تفاصيل alkaramdent.com/product/fast-pack/ انستغرام: https://alkaramdent.com/product/ultra-x/ https://instagram.com/alkaramdentalsupplies
message_without_#_length:	344
msg_is_ar:	1
msg_ar_len:	107
msg_is_en:	1
msg_en_len:	161
type:	photo
subattachments_count:	0
contains_main_topics:	0
distinct_emojis:	['🤩', '📣']
emoji_count:	2
unique_emoji_count:	2

توزع المنشورات النهائية بحسب النوع:

cover_photo:	14
profile_media:	6
photo:	944
album:	158
share:	22
video_inline:	298
video_direct_respons:	5

:text	9
:native_templates	139
:video_autoplay	4
:event	1
:video	30
:scheduled_live_video_post	1
:messenger_generic_template	1

٦,٢ أدوات البحث

تم استخدام لغة Python كلغة برمجية معتمدة في هذا البحث مع العديد من المكاتب والأدوات في معالجة البيانات ومن ثم التنفيذ العملي الفعلي نذكر أبرزها:

I. Facebook SDK:

هي عبارة عن مكتبة تساعد في التعامل مع واجهة برمجة التطبيقات Facebook Graph API المزودة من شركة ميتا وهي الطريقة الأساسية لتنفيذ عمليات المصادقة مع فيسبوك. [14]

II. Python Libraries:

- **Pandas**: هي مكتبة لتحليل البيانات والنمذجة. [13]
- **SciPy**: عبارة عن مجموعة من الحزم للرياضيات والعلوم والهندسة. [13] تم استخدامها في عملية دمج الميزات والتحويل إلى مصفوفة (hstack, csr_matrix).
- **Numpy**: هي مكتبة أساسية في الحوسبة العلمية تتيح التعامل مع المصفوفات المتعددة الأبعاد مع العديد من العمليات المنطقية والرياضية. [15]
- **Json**: هي عبارة عن مكتبة للتعامل مع البيانات التي تكون مكتوبة بصيغة json . [16] وتم استخدامها لأن البيانات العائدة من فيسبوك كانت بهذا التنسيق.
- **Re**: هي عبارة عن مكتبة تساعد في القيام بعمليات المطابقة مع التعبير الممرر. [16] وتم استخدامها للمطابقة مع تعبير الأحرف الإنكليزية والعربية من أجل إضافة ميزة تعبر عن لغة المنشور.

- **Collections**: تتيح هذه المكتبة القدرة على التعامل مع مجموعات البيانات بمختلف أشكالها بما فيها القواميس والقوائم وغيرها. [16]
- **Emoji**: هي عبارة عن مكتبة في بايثون تساعد على التعامل مع قائمة الرموز التعبيرية. [17]

III. Natural Language Toolkit (NLTK):

هي عبارة عن مجموعة من الأدوات تبلغ أكثر من 50 أداة تساعد في معالجة اللغات الطبيعية. [18]

بما في ذلك تقسيم الكلمات وترددها وإزالة كلمات التوقف أو الكلمات التي لا تحمل معنى وهذا ما تم استخدامه في معالجة نص المنشور واستخراج الكلمات الرئيسية (Stop words, Word tokenize and Word frequency).

IV. Scikit-learn:

مجموعة من الأدوات البسيطة والفعالة التي يتم استخدامها في تطبيقات تعلم الآلة وعمليات التنبؤ. [19] حيث تم استخدام الأدوات التالية: train_test_split، StandardScaler، OneHotEncoder، TfidfVectorizer، بالإضافة إلى:

- **Cosine Similarity:** حساب تشابه جيب التمام بين العينات في X و Y نقطياً، حيث أن X و Y تمثل المتجهين اللذين نبحث عن تشابه بينهما. [19]

$$K(X, Y) = \langle X, Y \rangle / (||X|| * ||Y||)$$

حيث:

X: يمثل متجه المنشور الجديد.

Y: يمثل متجه منشور تم نشره سابقاً.

$\langle X, Y \rangle$: حاصل الضرب النقطي بين المتجهين.

$||X||$: مقدار النورم الإقليدي Euclidean norm للمتجه X، وهو الجذر التربيعي لمجموع مربعات عناصر المتجه. يستخدم هذا النورم في حساب تشابه جيب التمام

لأنه يعطي مقياساً ثابتاً لطول المتجه، مما يساعد في حساب الزاوية بين المتجهات بدقة.

||Y||: مقدار النورم الإقليدي Euclidean norm للمتجه Y.

تم استخدام هذا القانون في استخراج المنشورات المشابهة للمنشور الجديد.

V. مقاييس التقييم

من أجل كل التجارب العملية سيتم تقييم الأداء من خلال المقاييس التالية:

- **MAE**: المتوسط الحسابي للأخطاء بين قيمة درجة انتشار المنشور PopScore الفعلية y_i ودرجة الانتشار المتوقعة $\hat{y_i}$ بالقيمة المطلقة.

$$MAE = \sum_{i=1}^{n} |y_i - \hat{y_i}|/n$$

حيث أن:

y_i: تمثل القيمة الحقيقية لدرجة انتشار المنشور.

$\hat{y_i}$: تمثل القيمة المتوقعة لدرجة انتشار المنشور.

n: عدد المنشورات الكلي.

- **MSE**: المتوسط الحسابي لمربعات الأخطاء بين درجة انتشار المنشور PopScore الفعلية y_i ودرجة الانتشار المتوقعة $\hat{y_i}$.

$$MSE = \sum_{i=1}^{n} (y_i - \hat{y_i})^2/n$$

- **RMSE**: الجذر التربيعي لقيمة المقياس السابق.

$$RMSE = \sqrt{\sum_{i=1}^{n} (y_i - \hat{y_i})^2 / n}$$

يتم استخدام المقاييس الثلاثة السابقة للمقارنة بين قيمة الانتشار المتوقعة PopScore بعد أن يتم تقييسها وبين القيمة الفعلية لها، وبما أن هذه المقاييس تمثل نسبة وجود الخطأ فإن القيمة المثلى لها هي الصفر.

- **F1 Score:** يتم استخدام هذا المقياس على قيمة تصنيف شعبية الانتشار Popularity لنحدد نسبة تطابق التصنيف المتوقع مع التصنيف الفعلي، وبما أنه يمثل دقة الصواب في التنبؤ فإن القيمة المثلى له هي 1.0.

$$F1\ Score = \frac{TP}{TP + \frac{FP + FN}{2}}$$

حيث:

TP (True Positives) : عدد المنشورات التي تم تصنيفها بشكل صحيح.

FP | FN (False Positives | False Negatives) : عدد المنشورات التي تم تصنيفها بشكل خاطئ.

٧. مقياس PopScore وطريقة حسابه

سنقوم في هذا القسم بشرح طريقة حساب المقياس PopScore بالاستفادة من المنشورات المشابهة وسلوك المستخدمين، ومن ثم كيفية إسناد تصنيف للمنشور يعبر عن شعبيته وانتشاره Popularity بالاعتماد على قيمة المقياس السابق PopScore لتحسين فعالية التسويق على وسائل التواصل الاجتماعي.

يبدأ التحليل بجمع البيانات حول أداء المنشورات السابقة، بما في ذلك عدد التفاعلات والتعليقات والمشاركات. يُمكن تحليل هذه البيانات لتحديد أنواع المحتوى التي حققت أفضل أداء، وتحديد الأوقات التي كان فيها التفاعل مع الجمهور أعلى. من خلال فهم اهتمامات الجمهور وتحليل سلوكهم، نستطيع تحديد أنواع المحتوى التي يتفاعلون معها بشكل أكبر، وكذلك الهاشتاغات المستخدمة والصفحات التي يتابعونها. بالإضافة إلى ذلك، يمكن اختبار أنواع مختلفة من المحتوى، مثل الصور ومقاطع الفيديو والنصوص والبث المباشر، لتحديد الأنواع التي تحقق أفضل تفاعل مع الجمهور.

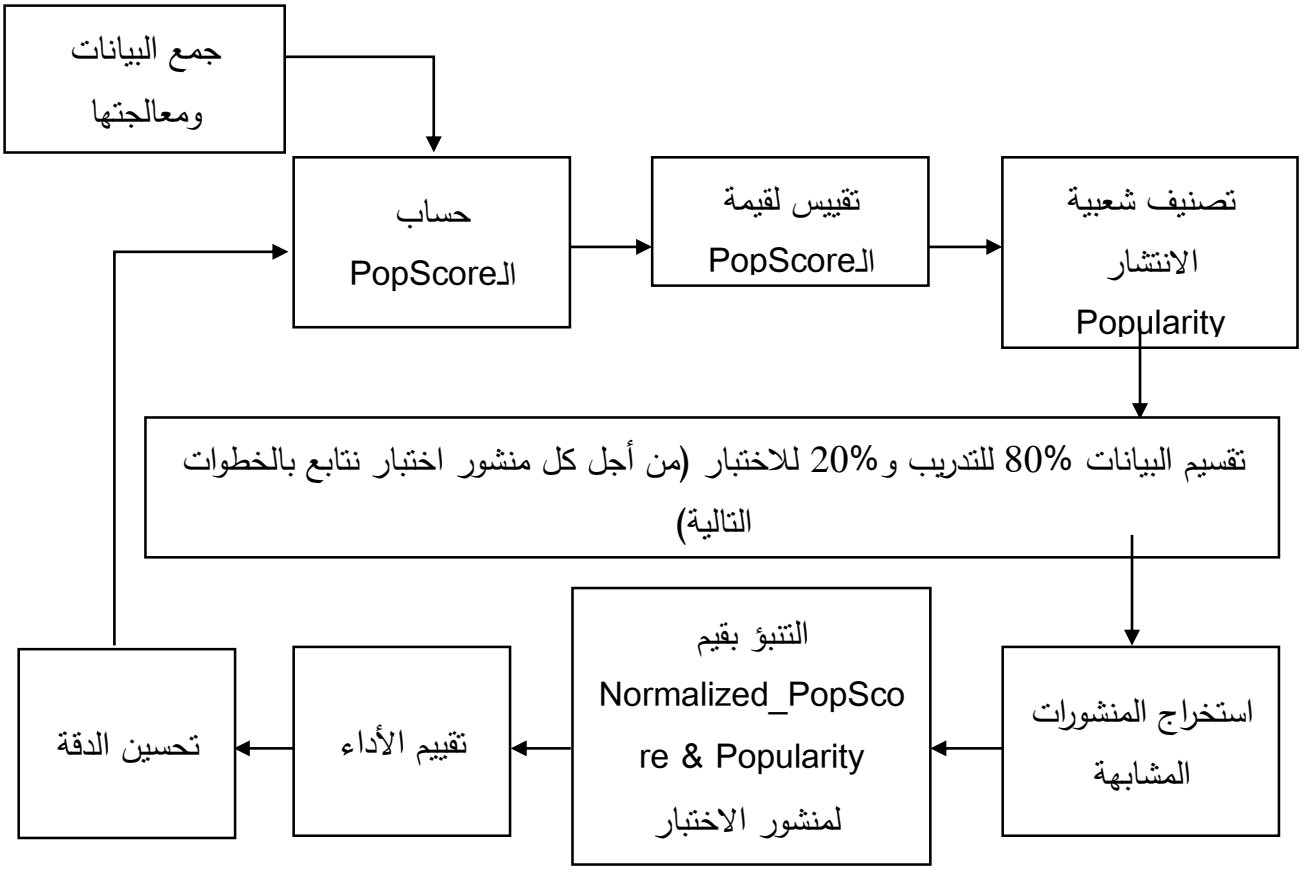
قمنا من خلال دراسة هذه الأفكار بتطوير طريقة تعتمد على مزيج من تقنيات تحليل البيانات ومعالجة النصوص لاستخراج الميزات الرئيسية للمنشورات وتحليل تشابهها. تم استخدام تقنية TF-IDF لاستخراج الميزات النصية من المنشورات، بالإضافة إلى تحليل تشابه جيب التمام Cosine Similarity لقياس مدى التشابه بين المنشورات المختلفة، وبذلك تم إيجاد مقياس جديد لتحديد درجة انتشار المنشور، اسميناه **PopScore**، بالإضافة إلى تصنيف فئوي يعبر عن شعبية المنشور بالاعتماد على درجة الانتشار، اسميناه **Popularity**.

مقياس PopScore هو عبارة عن مقياس جديد مصمم لتحديد درجة انتشار المنشور على وسائل التواصل الاجتماعي قبل نشره. يدمج هذا المقياس عدة متغيرات، تشمل عدد التعليقات والمشاركات والتفاعلات، مع أوزان مختلفة لكل منها لتحديد أي مزيج يعبر بدقة عن الانتشار الفعلي. من خلال الاستفادة من تقنيات تحليل البيانات ومعالجة النصوص في دراسة تفاعلات المستخدمين مع المنشورات المشابهة، يهدف PopScore إلى تقديم مقياس تنبؤي يساعد المسوقين في تقييم مدى نجاح وفعالية محتواهم.

بالإضافة إلى مقياس PopScore تم إضافة تصنيف Popularity وهو تصنيف فئوي يعبر عن شعبية المنشور بالاعتماد علي قيمة الانتشار PopScore التي تم حسابها مسبقاً.

يتيح هذا النهج فهماً أكثر دقة للعوامل التي تساهم في وصول المنشور وتفاعله، مما يعزز التخطيط الاستراتيجي لحملات التسويق عبر وسائل التواصل الاجتماعي.

يتم حساب PopScore و Popularity عبر مجموعة من الخطوات موضحة في المخطط التالي:



الشكل رقم (1): خطوات التنفيذ العملي.

سنوضح المخطط السابق من خلال كتابة كود زائف Pseudo Code:

```
# الخطوة 1: جمع المنشورات ومعالجتها
all_posts = collect_and_preprocess_posts()
# الخطوة 2: حساب درجة انتشار جميع المنشورات وتقييسها
for post in all_posts:
    post['Normalized_Score'] = calculate_popscore(post)
# الخطوة 3: تصنيف انتشار جميع المنشورات بحسب القيمة السابقة
for post in all_posts:
    post['Popularity'] = classify_popularity(post['Normalized_Score'])
# الخطوة 4: تقسيم المنشورات إلى تدريب واختبار
training_posts, testing_posts = split_data(all_posts, train_ratio=0.8,
test_ratio=0.2)
# الخطوة 5: من أجل كل منشور اختبار نستخرج المنشورات المشابهة له من مجموعة
التدريب ونقوم بعملية التنبؤ بناءً عليها
for test_post in testing_posts:
        similar_posts = extract_similar_posts(test_post,
        training_posts)      test_post['Normalized_PopScore'] =
        predict_normalized_popscore(test_post,
        similar_posts)
        test_post["Popularity"] =
        predict_popularity(test_post['Normalized_PopScore'])
#الخطوة 6: تقييم الأداء
mae = mean_absolute_error(true_scores, predicted_scores)
```

```
mse = mean_squared_error(true_scores, predicted_scores)
rmse = np.sqrt(mse)
f1 = f1_score(true_popularity, predicted_popularity,
average='weighted')
```

 # الخطوة 7: تحسين دقة المقياس

٧,١. خطوات التنفيذ:

١. **تحميل البيانات ومعالجتها:** قراءة البيانات ومعالجتها باستخدام الأدوات والمكتبات المذكورة سابقاً.
٢. **حساب الـPopScore:** إضافة عمود يعبر عن درجة اهتمام المستخدمين بكل منشور اسمه PopScore (سيتم مناقشة النسب المعتمدة في حساب هذه القيمة في ***الجدول رقم (2)***).
٣. **تقييس لقيمة الـPopScore:** القيام بعملية تقييس للعمود السابق وذلك بإضافة عمود آخر نسميه Normalized_score لجعل القيم ضمن مجال [0,1].
٤. **تصنيف شعبية الانتشار:** إضافة عمود تصنيف نصي للعمود السابق من خلال تحديد عدة عتبات نسميه Popularity (سيتم مناقشة العتبات المعتمدة في التصنيف في ***الجدول رقم (3)***).
٥. **تقسيم البيانات:** 20% للاختبار و80% للتدريب حيث أن عدد المنشورات الكلي 1632 تم أخذ 1305 منشوراً للتدريب و 327 منشوراً للاختبار.
٦. **استخراج المنشورات المشابهة:** من أجل كل منشور اختبار نحسب نسبة تشابهه مع كل منشورات التدريب.
٧. **التنبؤ بقيمة الـNormalized_score لمنشور الاختبار:** نحسب متوسط قيم Normalized_score للمنشورات المشابهة بدرجة كبيرة (نسبة التشابه أكبر أو

تساوي 0.7) وفي حال كان المنشور بمحتوى جديد ونسب التشابه لديه ضعيفة فإننا نأخذ متوسط القيم لكل المنشورات في مجموعة التدريب.

٨. **التنبؤ بتصنيف شعبية المنشور Popularity:** يتم اختيار التصنيف الأكثر تكراراً بين المنشورات المشابهة المستخرجة سابقاً.

٩. **تقييم الأداء:** في تنبؤ Popularity & Normalized_score.

١٠. **تحسين دقة المقياس:** بتغيير ميزات التشابه المدروسة ومعاملات حساب درجة الانتشار.

٧,٢. التجارب وتقييم الأداء:

سيناريوهات حساب درجة الـ PopScore:

بما أن هذا المقياس سيدمج عدة متغيرات، تشمل عدد التعليقات والمشاركات والتفاعلات، فنحن بحاجة إلى تحديد عدة سيناريوهات لتشكيل أوزان الأهمية للمتغيرات في حساب معادلة المقياس الجديد لتحديد أي مزيج يعبر بدقة عن الانتشار الفعلي، حيث في البداية قمنا بإعطاء أوزان متساوية للمتغيرات الثلاثة وبناءً على نتائج الخرج التي كنا نحصل عليها نجري تجارب على تغيير قيم هذه الأوزان، حيث تم إدراج جميع السيناريوهات التي تم تجربتها في ***الجدول رقم (2):***

الجدول رقم (2): السيناريوهات المطروحة لتشكيل أوزان المتغيرات في معادلة حساب مقياس الـ PopScore.

السيناريو	المعادلة
1	PopScore = (0.33 * reactions) + (0.33 * shares) + (0.33 * comments)
2	PopScore = (0.40 * reactions) + (0.20 * shares) + (0.40 * comments)

PopScore = (0.60 * reactions) + (0.10 * shares) + (0.30 * comments)	3
PopScore = (0.80 * reactions) + (0.05 * shares) + (0.15 * comments)	4
PopScore = (0.35 * reactions) + (0.05 * shares) + (0.60 * comments)	5
PopScore = (0.15 * reactions) + (0.05 * shares) + (0.80 * comments)	6
PopScore = (0.05 * reactions) + (0.05 * shares) + (0.90 * comments)	7
PopScore = (0.0 * reactions) + (0.0 * shares) + (1.0 * comments)	8

بعد حساب قيمة الـPopScore يتم تطبيق عملية التقييس عليها Normalization لنحصل على قيم تتراوح بين [0,1] ونسمي القيمة Normalized_score وهي التي يتم اعتمادها في المراحل التالية في تنبؤ تصنيف الانتشار وتقييم الأداء.

عتبات تصنيف شعبية المنشور Popularity (هو عبارة عن تصنيف لدرجة الـ Normalized_score):

بعد الحصول على قيمة انتشار المنشور PopScore كقيمة رقمية بين الـ[0,1] نريد إضافة تصنيف فئوي نسميه Popularity يعبر عن هذه القيمة كمقياس إضافي يدخل في عملية تقييم الأداء العام، حيث يتم وضع عدة عتبات وبناءً على تجاوز القيمة لعتبة معينة يتم إضافة التصنيف المناسب، في ***الجدول رقم (3)*** تم توضيح العتبات المطبقة في عملية التصنيف هذه:

الجدول رقم (3): العتبات المطبقة في إضافة تصنيف Popularity لقيمة المقياس PopScore.

التصنيف	مجال العتبات
Popularity = Bad	Normalized_score < 0.4
Popularity = Good	0.4 <= Normalized_score < 0.7
Popularity = Very Good	Normalized_score >= 0.7

تم تحديد هذه العتبات بناءً على نهج عملي بسيط يهدف إلى تسهيل عملية تصنيف المنشورات إلى فئات مفهومة ومباشرة تعكس الأداء العام للمحتوى من خلال تقديم تصنيف مبدئي واضح يميز بين المنشورات ذات الأداء المنخفض "سيء" والمتوسط "جيد" والعالي "جيد جداً"، وهو تقسيم يساهم في توفير إطار أولي لتحليل الأداء، كما أنه يمكن اعتبار هذه العتبات مرنة وقابلة للتعديل مستقبلاً بناءً على طبيعة البيانات المدروسة والأهداف البحثية.

السيناريوهات المطروحة لإدخال الميزات في استخراج المنشورات المشابهة:

من أجل استخراج المنشورات المشابهة يجب تحديد الميزات التي على أساسها يتم حساب التشابه، تم تجربة إدخال عدة ميزات حيث أن بعضها أثر بشكل إيجابي على الأداء مثل تضمين الهاشتاغات في نص المنشور، طول المنشور، اللغة المستخدمة، نوع المنشور، عدد الوسائط المضمنة فيه وهل يحتوي المنشور على كلمات رئيسية أم لا، وبعضها الآخر أثر بشكل سلبي مثل دراسة أشكال الرموز التعبيرية المستخدمة، **في *الجدول رقم (٤)*** تم تحديد السيناريوهات المطروحة لإدخال الميزات في عملية حساب التشابه بين المنشورات.

الجدول رقم (4): السيناريوهات المطروحة لإدخال الميزات في عملية حساب التشابه بين المنشورات.

السيناريو	ميزات المنشور
1	نص المنشور بالكامل مع الهاشتاغات نص المنشور الأساسي بدون الهاشتاغات طول نص المنشور الأساسي بدون الهاشتاغات اللغة المستخدمة في المنشور طول المحارف العربية طول المحارف الإنكليزية عدد الوسائط المستخدمة في المنشور (5 صور) هل المنشور يحوي على الكلمات الرئيسية عدد الرموز التعبيرية (بعض الرموز يتم استخدامها أكثر من مرة ضمن نفس المنشور) عدد الرموز التعبيرية الفريدة نوع المنشور: (صورة، فيديو، ألبوم ..الخ) الرموز التعبيرية في نص المنشور 😁😂👍
2	نفس الميزات السابقة باستثناء: نص المنشور الأساسي بدون الهاشتاغات الرموز التعبيرية في نص المنشور 😁😁👍

٧,٣. النتائج والمناقشة:

من أجل كل تجارب التنفيذ تم حساب قيمة Normalized_score بتطبيق السيناريوهات الثمانية كاملة في حساب معادلة الـPopScore.

التجارب المنفّذة:

تم تنفيذ تجربتين عمليتين، حيث تم حساب درجة الـPopScore من أجل كل سيناريوهات ***الجدول رقم (2)*** ليصبح لدينا القيم التالية:

PopScore1, PopScore2, PopScore3…PopScore8

ومن ثم قمنا بتقييس لهذه القيم لتصبح ضمن المجال [0,1]:

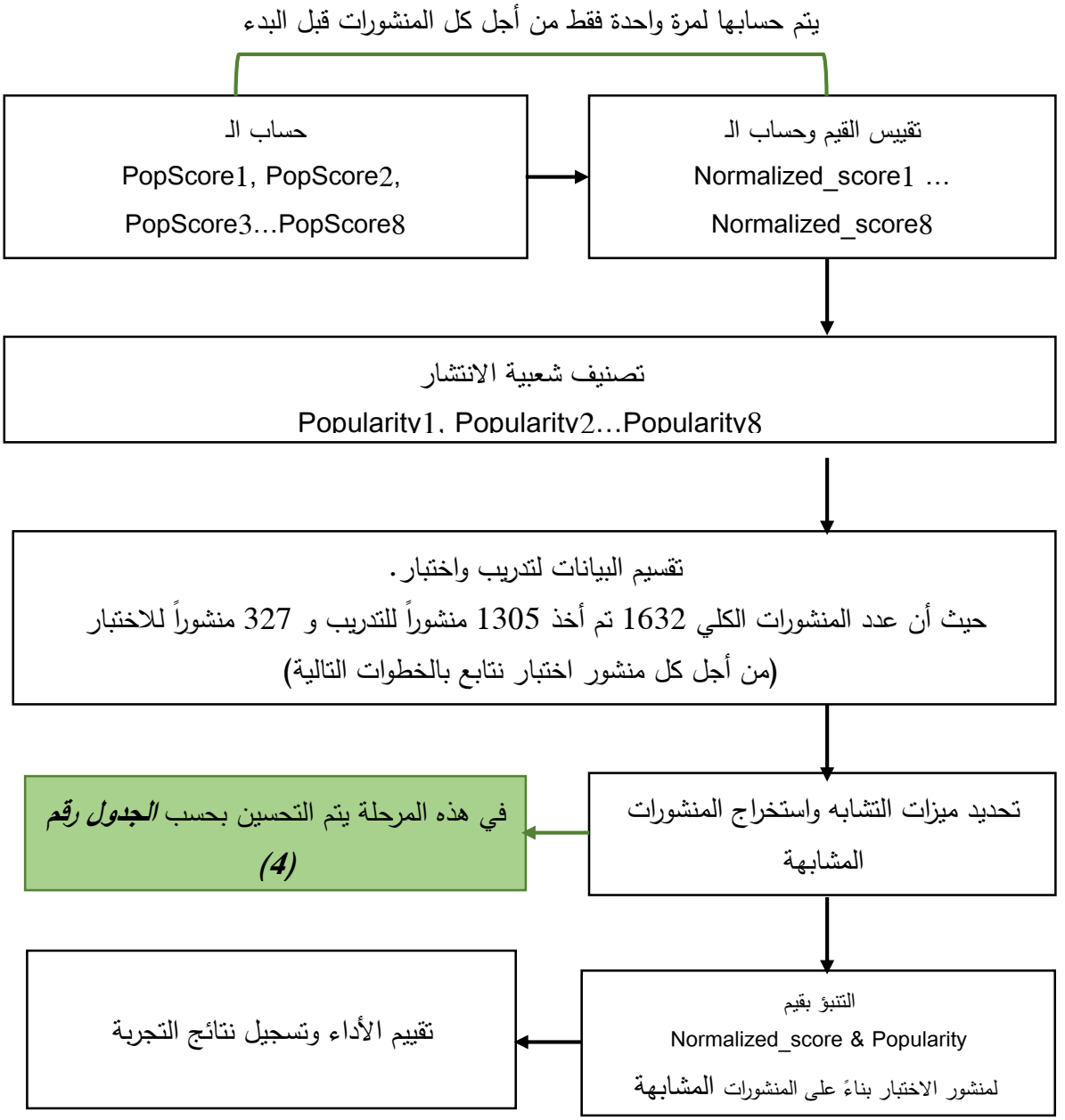
Normalized_score1, Normalized_score2,

Normalized_score3…Normalized_score8

تشترك التجربتان بأنه تم حساب مقاييس الأداء من أجل كل سيناريوهات القيم المقيسّة Normalized_score السابقة لمعرفة أي الأوزان أعطت دقة أعلى، وتختلف عن بعضها بتحديد ميزات المنشور التي تدخل في عملية استخراج المنشورات المشابهة.

في التجربة الأولى تم حساب Normalized_score من أجل السيناريوهات الثمانية جميعها وتطبيق عتبات التصنيف الموضحة في ***الجدول رقم (٣)*** والتي تمثل ثلاث تصنيفات للشعبية (Bad, Good, Very Good) مع إدخال ميزات المنشور الموجودة في السيناريو الأول من الجدول ***الجدول رقم (٤)*** والتي تأخذ بعين الاعتبار نص المنشور بالكامل، نص المنشور الأساسي بعد إزالة الهاشتاغات وطوله، اللغة المستخدمة في المنشور وطول المحارف العربية والإنكليزية فيها، موضوع المنشور وكلماته الأساسية، نوع الوسائط المستخدمة فيه وعددها، عدد الرموز التعبيرية المستخدمة وأشكالها. قمنا بتنفيذ التجربة وتسجيل النتائج ومن ثم بدأت عملية التحسين تم افتراض أن ميزات المنشور المدروسة في عملية التحليل قد تؤثر في دقة المقياس لذلك تم تجربة إزالة كل ميزة بشكل منفصل لنرى تأثير كل منها إلى أن وجدنا أن جميع الميزات المدروسة في التجربة الأولى لها تأثير إيجابي في تحسين دقة المقياس باستثناء ميزة نص المنشور الأساسي بدون الهاشتاغات وميزة أشكال الرموز التعبيرية المستخدمة لهما تأثير سلبي بسيط لذلك تم استبعادهم في تنفيذ **التجربة الثانية** واعتماد السيناريو الثاني من الميزات الموجودة في ***الجدول رقم (٤)***، والتي أعطت نتائجاً أفضل.

كلا التجربتين تمران بمخطط التنفيذ الموجود في ***الشكل رقم (1)***، قمنا بالتعديل عليه ليتم تحديد مكان إضافة التحسينات فحصلنا على ***الشكل رقم (٢)*** والذي يحدد المراحل التي تم فيها تنفيذ التجارب وتحسين دقة المقياس:



الشكل رقم (2): خطوات تنفيذ التجارب والتحسين.

نتائج التجربة الأولى:

بعد أن قمنا بجمع البيانات ومعالجتها لتصبح جاهزة للتحليل تم اتباع ***منهجية التنفيذ الموضحة في الشكل رقم (2)*** حيث تم إضافة 8 أعمدة تمثل كافة سيناريوهات الأوزان للمتغيرات التي تدخل في مقياس الانتشار (PopScore1 ... PopScore8) والموضحة ***في الجدول رقم (2)***، وتم إضافة 8 أعمدة لتصنيف القيم السابقة بالتصنيفات التالية: كل ماهو أقل من 0.4 أي ضمن المجال [0,0.4[يُصنف على أنه سيء Bad، والمجال [0.4,0.7[يُصنف على أنه جيد الشعبية Good، وأخيراً المجال [0.7,1.0] يُصنف على أنه أفضل أنواع الانتشار والشعبية Very Good. وتم تسمية الأعمدة على التتالي: (Popularity1…Popularity8) ومن ثم تم تقسيم البيانات إلى تدريب واختبار حيث أن عدد المنشورات الكلي 1632 تم أخذ 1305 منشوراً للتدريب و 327 منشوراً للاختبار، ثم المرور على كل منشور اختبار واستخراج كافة المنشورات المشابهة له من بيانات التدريب حيث تم اعتماد مزيج من الميزات النصية والعددية والتصنيفات الفئوية في حساب عملية التشابه، والمتمثلة بميزات ***السيناريو الأول من الجدول رقم (٤)***، وقد تم استخدام TfidfVectorizer لمعالجة ميزة نص المنشور الكلي مع تضمين الهاشتاغات الموجودة فيه وميزة نص المنشور الأساسي حيث أنه يقوم بتحويل السلسلة النصية إلى متجهات عددية، كما تم استخدام OneHotEncoder لتحويل الميزات الفئوية إلى رقمية وقد تم إنشاء قاموس بالرموز التعبيرية لتحويلها إلى تمثيل عددي، وأخيراً الميزات العددية: طول المنشور، عدد المحارف العربية والإنكليزية، الكلمات الرئيسية المضمنة، نوع المنشور، وعدد الوسائط الموجودة فيه، عدد الرموز التعبيرية ككل وعددها الفريد بدون تكرار، حيث تم دمج جميع الميزات السابقة والقيام بعملية تقييس StandardScaler لنضمن أن جميع الميزات لها نفس النوع، ثم تم استخدام مقياس تشابه جيب التمام Cosine Similarity في حساب درجة التشابه بين المنشور الجديد والمنشورات السابقة بعد أن تم تحويل المنشورات إلى متجهات رقمية كل عنصر في المتجه يمثل تكرار كلمة أو مصطلح في النص، حيث أن درجة التشابه المساوية لـ 0 تعني عدم وجود تشابه نهائياً، ودرجة التشابه المساوية لـ 1 تعني تطابق تام في المحتوى.

بعد حساب درجة التشابه مع كل المنشورات السابقة، يتم اعتماد المنشورات التي نسبة تشابهها مع المنشور الجديد أكبر من 0.7 على أنها منشورات مشابهة ويكون حاصل متوسط قيم Normalized_score لهذه المنشورات هو قيمة انتشار المنشور الجديد وتصنيف Popularity الأكثر تكراراً في هذه المنشورات هو تصنيف شعبية الانتشار للمنشور الجديد:

لنفترض أن S_i يمثل درجة تشابه المنشور الجديد مع المنشور السابق i وعدد المنشورات المشابهة يرمز له n وبالتالي المنشورات المشابهة هي التي يكون فيها . $S_i \geq 0.7$

$$\text{Predicted_normalized_score} = \frac{\sum_{i=1}^{n} Normalized_score_i}{n}$$

لنفترض أن C_i يمثل تصنيف Popularity للمنشور i وتصنيف شعبية انتشار المنشور الجديد سيكون التصنيف الأكثر تكراراً بين المنشورات المشابهة.

$$\text{Predicted_popularity} = \text{Mode}(\{C_i \mid S_i \geq 0.7\})$$

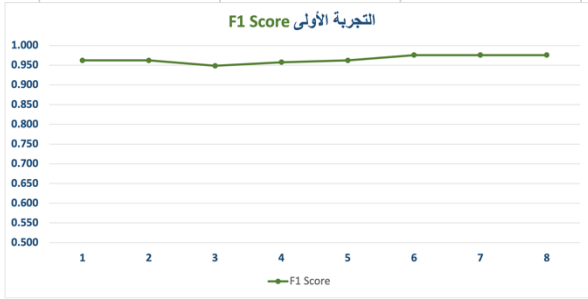
قد يخطر في البال تساؤلاً عن حالة عدم وجود منشورات مشابهة بنسبة 0.7 أو أكثر، في هذه الحالة يتم تضمين كافة المنشورات السابقة في حساب التنبؤ.

بعد حساب التنبؤ يتم قياس أداء المقياس وتقييمه من خلال مقياس MAE وهو يعبر عن المتوسط الحسابي للخطأ بالقيمة المطلقة في تنبؤ درجة الانتشار Predicted_normalized_score، ومقياس MSE الذي يعبر عن المتوسط الحسابي لمربعات الأخطاء ومقياس RMSE وهو جذر المقياس السابق ومن أجل أفضل قيمة للمقاييس السابقة يجب أن تكون أقرب مايمكن إلى الصفر، أما بالنسبة لدقة التنبؤ بتصنيف شعبية الانتشار Predicted_popularity تم استخدام مقياس F1 Score وكلما كانت

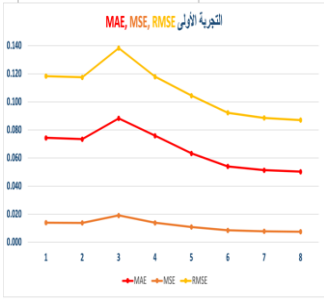
قيمته أعلى كانت الدقة أفضل، حيث أن نتائج التجربة الأولى موضحة في ***الجدول رقم (5):***

الجدول رقم (5): مقاييس الأداء في التجربة الأولى.

التجربة الأولى				
Normalized Score	MAE	MSE	RMSE	F1 Score
1	0.074	0.014	0.118	0.962
2	0.073	0.014	0.118	0.962
3	0.088	0.019	0.138	0.948
4	0.076	0.014	0.118	0.957
5	0.063	0.011	0.104	0.962
6	0.054	0.009	0.092	0.976
7	0.051	0.008	0.089	0.976
8	0.050	0.008	0.087	0.976



الشكل رقم (3): مقياس تصنيف الشعبية للتجربة الأولى.



الشكل رقم (4): مقاييس درجة الانتشار للتجربة الأولى.

العمود الأول في ***الجدول رقم (5)*** يمثل درجة الانتشار المقيسّة Normalized الناتجة عن المعادلة المستخدمة في توزيع الأوزان في عملية حساب درجة الانتشار PopScore ، وهي عبارة عن 8 سيناريوهات موضحة في ***الجدول رقم (2)***، حيث أن السيناريوهات في البداية أعطت أوزان متساوية للمتغيرات الثلاثة ثم جعلت الوزن الأقل هو عدد المشاركات، ثم

رفعت وزن التفاعلات على حساب الأوزان الأخرى، ثم رفعت وزن التعليقات على حساب الأوزان الأخرى على الترتيب، أما بالنسبة للأعمدة الأخرى في الجدول فهي تمثل دقة المقياس من أجل كل معادلة.

عندما قمنا بتمثيل هذه النتائج على مخططات بيانية ظهرت لدينا الأشكال رقم ***(3)*** و ***(4)***، نجد أن مقاييس الخطأ الثلاثة تعطي نفس الأداء في التحسن أو التراجع وكذلك مقياس F1 Score أي أنه لا يوجد لدينا تعارض بين هذه المقاييس، تم البدء بمعادلة متساوية الأوزان للمتغيرات الثلاثة:

PopScore_1 = (0.33 * reactions) + (0.33 * shares) + (0.33 * comments)

ثم تم تخفيض وزن المشاركات قليلاً مع تساوي وزن التفاعلات والتعليقات فتحسنت المقاييس الثلاثة:

PopScore_2 = (0.40 * reactions) + (0.20 * shares) + (0.40 * comments)

نلاحظ أن جميع المقاييس كان الأداء الأسوأ لها في السيناريو الثالث والذي يعطي الوزن الأكبر لعدد التفاعلات ويخفض من وزن التعليقات عن السابق:

PopScore_3 = (0.60 * reactions) + (0.10 * shares) + (0.30 * comments)

تمت المتابعة بزيادة وزن التفاعلات وتخفيف وزن المشاركات حتى الحالة الرابعة وتم تحسين القيم، ثم تم تجربة عدة سيناريوهات كان فيها وزن المشاركات هو الأخف ووزن التفاعلات هو الأوسط ووزن التعليقات هو الأعلى بدأ المنحني في ***الشكل رقم (٤)*** بالاقتراب من الصفر بشكل ملحوظ أي بتحسن المقاييس ومع استمرار تخفيض وزن المشاركات وزيادة وزن التعليقات نجد أن قيم المقاييس من الحالة 6 إلى الحالة 8 شبه متطابقة:

PopScore_6 = (0.15 * reactions) + (0.05 * shares) + (0.80 * comments)

PopScore_8 = (0.0 * reactions) + (0.0 * shares) + (1.0 * comments)

MAE_6 = 0.054 , MAE_8 = 0.050

أما بالنسبة لمقياس F1 Score في ***الشكل رقم (3)*** نجد أن دقة التنبؤ قد ثبتت عند الحالة 6 إلى الحالة 8 بقيمة مقبولة جداً وتبلغ:

F1_Score_6 = 0.976 , F1_Score_8 = 0.976

عند تحليل النتائج السابقة، نجد أنه عندما كانت الأوزان متساوية بين التفاعلات، المشاركات والتعليقات حقق النموذج أداءً مقبولاً، إلا أن قيم الخطأ كانت أعلى نسبياً هذا يشير إلى أن تأثير التوزيع المتساوي للأوزان لم يكن الخيار الأمثل وهذا أمر منطقي، حيث أن التوزيع المتساوي لن يعكس الأهمية النسبية لكل عامل في تحديد درجة انتشار المنشور. وعند تعديل الأوزان من خلال رفع وزن التفاعلات والتعليقات قليلاً وتخفيض وزن المشاركات كمحاولة لتحسين الأداء، لوحظ تحسن في المقاييس الثلاثة وبدأت مؤشرات الخطأ بالتناقص، هذا التعديل يعكس فكرة أن التعليقات والتفاعلات تمثل عادةً مؤشرات أكثر دقة على شعبية المنشور مقارنةً بعدد المشاركات.

ومع ذلك، عند الاستمرار بزيادة وزن التفاعلات على حساب التعليقات لاحظنا أن أداء النموذج بدأ بالتراجع، حيث انخفضت دقة F1 وازدادت قيم الخطأ، مما يشير إلى أن زيادة وزن التفاعلات على حساب التعليقات ليس استراتيجية فعّالة. وهذا يُعزى إلى أن التفاعلات وحدها دون الاهتمام الكافي بالتعليقات لا تعكس الاهتمام الحقيقي من الجمهور. فالتعليقات تعبّر عن تفاعل أعمق مقارنة بالتفاعلات البسيطة.

تم الوصول إلى أفضل أداء عند تخصيص وزن عالٍ للتعليقات وتخفيض وزن التفاعلات والمشاركات، حيث تميزت المقاييس بأقل قيم خطأ ممكنة وحققت F1 Score مرتفع عند ٠,٩٧٦. هذا لأن التعليقات تُعتبر مؤشراً أقوى لمدى تفاعل الجمهور مع المحتوى، حيث تتطلب من المستخدمين وقتاً وجهداً أكبر للتفاعل مقارنة بالتفاعلات أو المشاركات. وهذا يؤكد على أهمية التعليقات كعامل رئيسي للتنبؤ بدرجة انتشار المنشور.

في هذه التجربة وصلنا إلى أداء يعتبر جيد جداً، حيث أن مقاييس الخطأ كانت أقل من الواحد وأقرب إلى الصفر بنسبة مقبولة كما أننا وصلنا إلى قيمة عالية جداً في مقياس F1 Score.

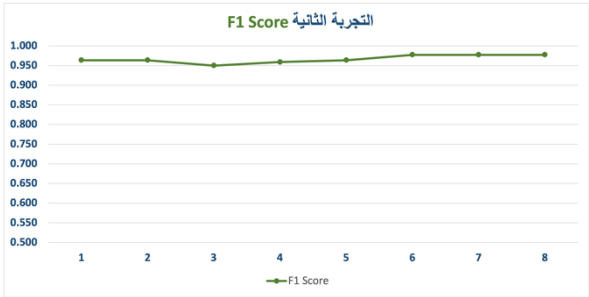
التجربة الثانية:

في التجربة الأولى قمنا بدراسة العديد من الميزات في حساب عملية التشابه بين المنشورات، لنفترض أن الميزات المدروسة قد تؤثر على أداء المقياس، انطلاقاً من النتيجة التي حصلنا عليها في التجربة الأولى تمت إزالة ميزة تلو الأخرى من عملية حساب التشابه بين المنشورات ومن ثم قمنا بقياس الدقة لمعرفة أي الميزات أثرت بشكل إيجابي أو سلبي على الأداء، إلى أن وجدنا أن دراسة نص المنشور بالكامل مع الهاشتاغات المضمنة فيه مع دراسة نص المنشور الأساسي كميزتين معتمدتين معاً أثر بشكل سلبي على الأداء، لذلك تم استبعاد ميزة نص المنشور الأساسي من عملية حساب التشابه حيث أن دمج الهاشتاغات مع النص ممكن أن يساهم في المعنى العام للمنشور، لذلك تم تضمين ميزة نص المنشور بالكامل فقط دون ميزة النص الأساسي وقد أدى ذلك إلى أداءٍ أفضل، كما أننا لاحظنا بأن دراسة أشكال الرموز التعبيرية المستخدمة أثر بشكل سلبي على أداء المقياس لذلك تم استبعادها من حساب عملية التشابه، وفي النهاية تم الإبقاء على الميزات المذكورة في السيناريو الثاني من ***الجدول رقم (4)***.

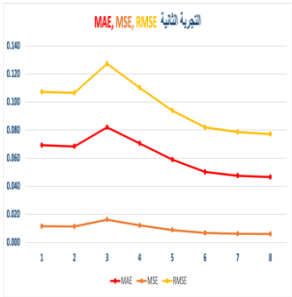
يظهر ***الجدول رقم (6)*** نتائج التجربة الثانية بعد التعديل على ميزات التشابه المدروسة في التجربة الأولى، كما أن الأشكال رقم ***(5)*** و ***(6)*** تظهر الشكل البياني لأداء المقياس.

التجربة الثانية				
Normalized Score	MAE	MSE	RMSE	F1 Score
1	0.069	0.011	0.107	0.963
2	0.068	0.011	0.107	0.963
3	0.082	0.016	0.127	0.950
4	0.070	0.012	0.110	0.959
5	0.059	0.009	0.094	0.963
6	0.050	0.007	0.082	0.977
7	0.048	0.006	0.079	0.977
8	0.047	0.006	0.077	0.977

الجدول رقم (6): مقاييس الأداء في التجربة الثانية.



الشكل رقم (5): مقياس تصنيف الشعبية للتجربة الثانية.



الشكل رقم (6): مقاييس درجة الانتشار للتجربة الثانية.

نلاحظ أن شكل الخط البياني لم يتغير عن التجربة الأولى ولكن القيم تحسنت عن السابق.

الجدول رقم (7): نسبة التحسين بين التجربة الأولى والثانية.

المقياس	التجربة الأولى	التجربة الثانية	التحسين
MAE_8	0.050	0.047	0.003
MSE_8	0.008	0.006	0.002
RMSE_8	0.087	0.077	0.010
F1_Score_8	0.976	0.977	0.001

بناءً على النتائج التي تم تحقيقها، يمكن تفسير التحسن الطفيف في الأداء بعد تعديل الميزات المستخدمة بأن الميزات المستبعدة، مثل نص المنشور الأساسي والرموز التعبيرية، لم تكن ذات تأثير إيجابي كبير على عملية حساب التشابه، وبالتالي فإن استبعاد هذه الميزات أدى إلى تحسين دقة التنبؤ بدرجة الانتشار وتقليل الأخطاء في المقاييس المستخدمة. ويعود ذلك إلى أن النص الكامل مع الهاشتاغات يوفر سياقاً أكثر شمولية ودقة لمعنى المنشور، بينما الرموز التعبيرية قد لا تضيف قيمة تحليلية كافية. وهذا يؤكد أهمية اختيار الميزات ذات العلاقة المباشرة بالأداء لتحقيق نتائج أفضل.

على الرغم من وجود تحسن بفارق بسيط في القيم عن التجربة الأولى عند تغيير الميزات المستخدمة في استخراج التشابه، يمكن أن نعتبر أننا وصلنا إلى قيم مقبولة جداً في مقاييس حساب الخطأ لدرجة الانتشار Predicted_Popscore، كما أننا وصلنا إلى أفضل قيمة يمكن الوصول إليها في مقياس التصنيف الفئوي لشعبية المنشور Predicted_Popularity.

تناولنا في هذا القسم كيف تم تطوير مقياس جديد لتقييم نجاح المنشورات قبل نشرها، وكيف تم دمج مفهوم فلترة المحتوى Content Filtering في عملية التنبؤ الحسابية والتصنيفية، حيث تم تنفيذ تجربتين عمليتين بغرض تحسين الأداء بطريقة تراكمية تزايدية بالاعتماد على نتائج التجربة السابقة يتم تنفيذ التجربة التالية، وتبين لنا أن عدد التعليقات يعتبر المتغير

الأهم في عملية التنبؤ بدرجة الانتشار مقابل عدد التفاعلات والمشاركات، كما أن وجود الهاشتاغات ضمن نص المنشور يساهم في تحسين عملية استخراج التشابه وبالتالي تحسين الأداء حيث أنها ممكن أن تساهم في المعنى العام للمنشور، كما أن وجود الرموز التعبيرية في محتوى المنشور ساهم بتحسين بسيط جداً يكاد أن يكون غير ملحوظ، ولكن دراسة أشكال الرموز التعبيرية بحد ذاتها كان لها تأثير سلبي وبالتالي يمكن القول أنه يمكن أن يكون لوجود الرموز التعبيرية في نص المنشور أثراً في الانتشار لكن لا علاقة لشكل الرموز المستخدمة بذلك، أي لا يوجد رمز يساهم في الانتشار أكثر من رمز آخر، وبذلك تم الوصول إلى دقة تنبؤ وصلت إلى 0.977.

٨. الخاتمة والآفاق المستقبلية

يُقدم هذا البحث خطوة هامة في مجال تحليل البيانات التسويقية على وسائل التواصل الاجتماعي. فقد تمكنا من تطوير مقياس جديد يُتيح للمسوقين توقع مدى نجاح منشوراتهم على فيسبوك قبل نشرها، وذلك من خلال تحليل المنشورات المشابهة وسلوك المستخدمين.

نطمح في المستقبل إلى تطوير هذا البحث بعدة اتجاهات وأهمها **إمكانية تقديم توصيات:** نهدف إلى تطوير منهجية تُقدم للمسوقين توصيات مخصصة لتحسين محتوى منشوراتهم وزيادة تفاعل المستخدمين معها. سيتضمن ذلك تحليل العوامل التي تؤثر على نجاح المنشور، مثل نوع المحتوى والصور والتوقيت والعلامات التجارية.

٩. المراجع

[1] JAGONGO A, 2013 - The Social Media and Entrepreneurship Growth (A New Business Communication Paradigm among SMEs in Nairobi). International journal of humanities and social science, p. 15, Nairobi, 15.

[2] AYAD N, SHIRKHODAIE M, SULAIMAN A, 2022 - The growth of e-marketing under covid-19 social distancing: analytical descriptive study in alzaidy commercial company for electronic marketing. International journal of research in social sciences & humanities, p. 20, Iraq, 20.

[3] MASON A, NARCUM J, MASON K, AWAN U, 2021 - Social media marketing gains importance after Covid-19. Cogent Business & Management, p. 8.

[4] ALMURTADHA Y, 2018 - Mining trending hash tags for Arabic sentiment analysis. (IJACSA) International Journal of Advanced Computer Science and Applications, p. 6, Kingdom of Saudi Arabia, 6.

[5] CHAULAGAIN S, MAHARJAN S, 2020 - Social Media as a Strategic Marketing Tool to Study the Consumer Behavior of Coca-Cola by K-Means Cluster Analysis: Evidence from Facebook. Proceedings of 8th IOE Graduate Conference, p. 7, Nepal, 7.

[6] JACOBSON J, GRUZD A, 2020 - Social media marketing: Who is watching the watchers? Journal of Retailing and Consumer Services, p. 29, USA, 29.

[7] XU Z, VAIL C, KOHLI AS, TAJDINI S, 2021 – Understanding changes in a brand's core positioning and customer engagement: a sentiment analysis of a brand-owned Facebook site. Journal of Marketing Analytics, p. 14, USA, 14.

[8] DAI Y, WANG T, 2021 – Prediction of customer engagement behaviour response to marketing posts based on machine learning. Unpublished, p. 21, China, 21.

[9] FLETCHER K-A, GBADAMOSI A, 2022 – Examining social media live stream's influence on the consumer decision making: a thematic analysis. Electronic Commerce Research, p. 31, UK, 31.

[10] BENSLAMA T, JALLOULI R, 2020 – Clustering of Social Media Data and Marketing Decisions. Springer International Publishing, p. 13, Switzerland, 13.

[11] AYDIN G, URAY N, SILAHTAROGLU G, 2021 – How to Engage Consumers through Effective Social Media Use—Guidelines for Consumer Goods Companies from an Emerging Market. Journal of Theoretical and Applied Electronic Commerce Research, p. 23, Turkey, 23.

[12] META Platform, "Meta Platforms, Inc.," Meta Company, 2024. [Online]. Available: https://about.meta.com/.

[13] PYTHON Website, "Python Official Website," Python, 2024. [Online]. Available: https://www.python.org/.

[14] FACEBOOK SDK, "Facebook SDK for Python," Facebook SDK for Python, 2024. [Online]. Available: https://facebook-sdk.readthedocs.io/en/latest/.

[15] NUMPY Website, "Numpy Official Website," Numpy, 2024. [Online]. Available: https://numpy.org/devdocs/.

[16] PYTHON Documentation, "Python Documentation," Python, 2024. [Online]. Available: https://docs.python.org/3/library/.

[17] EMOJI, "Emoji for Python," Emoji for Python, 2024. [Online]. Available: https://pypi.org/project/emoji/.

[18] NLTK, "Natural Language Toolkit," Natural Language Toolkit, 2024. [Online]. Available: https://www.nltk.org/.

[19] SCIKIT-LEARN, "Scikit-learn, Machine Learning in Python," Scikit-learn, 2024. [Online]. Available: https://scikit-learn.org/stable/index.html.