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.

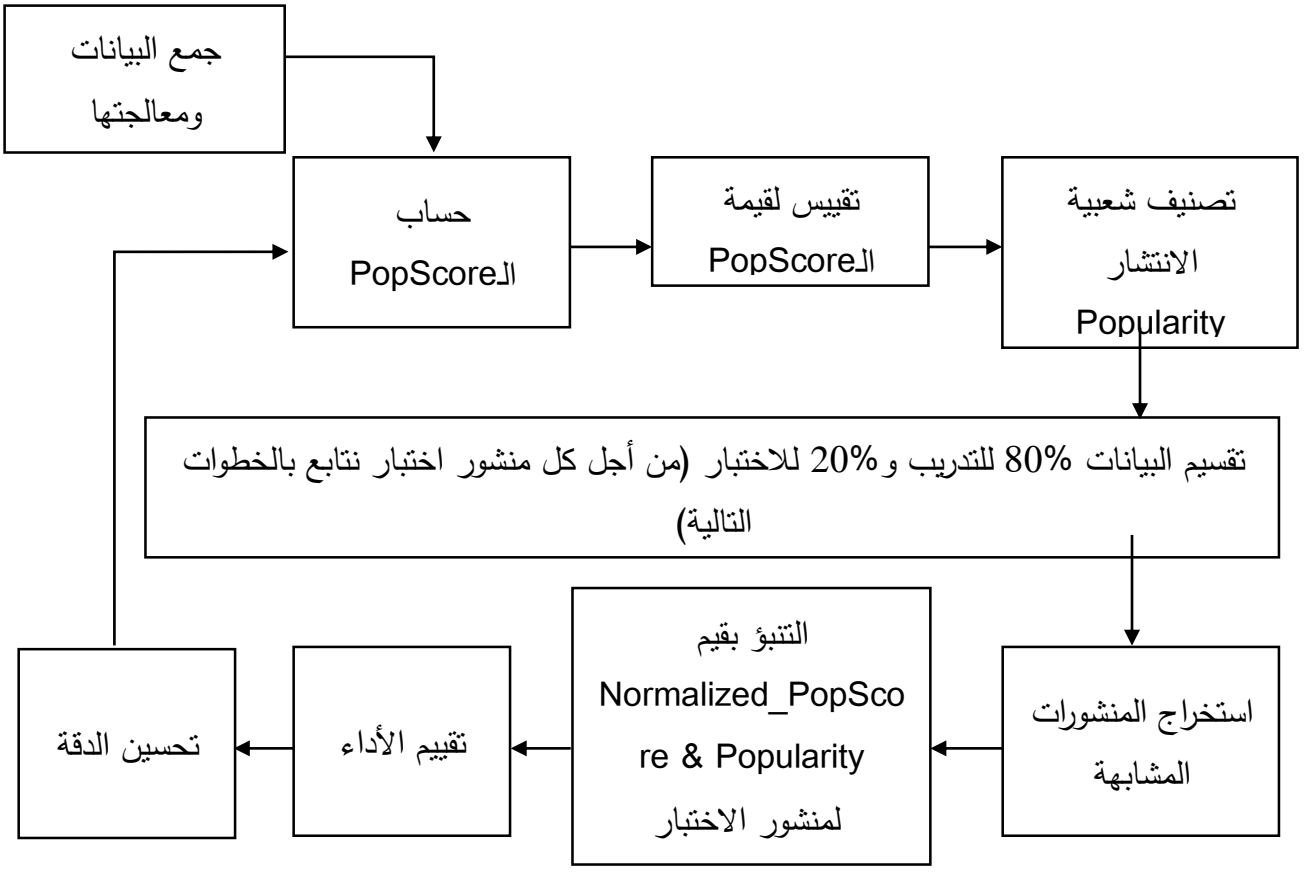
قمنا من خلال دراسة هذه الأفكار بتطوير طريقة تعتمد على مزيج من تقنيات تحليل البيانات ومعالجة النصوص لاستخراج الميزات الرئيسية للمنشورات وتحليل تشابهها. تم استخدام تقنية TF-IDF لاستخراج الميزات النصية من المنشورات، بالإضافة إلى تحليل تشابه جيب التمام Cosine Similarity لقياس مدى التشابه بين المنشورات المختلفة، 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]:

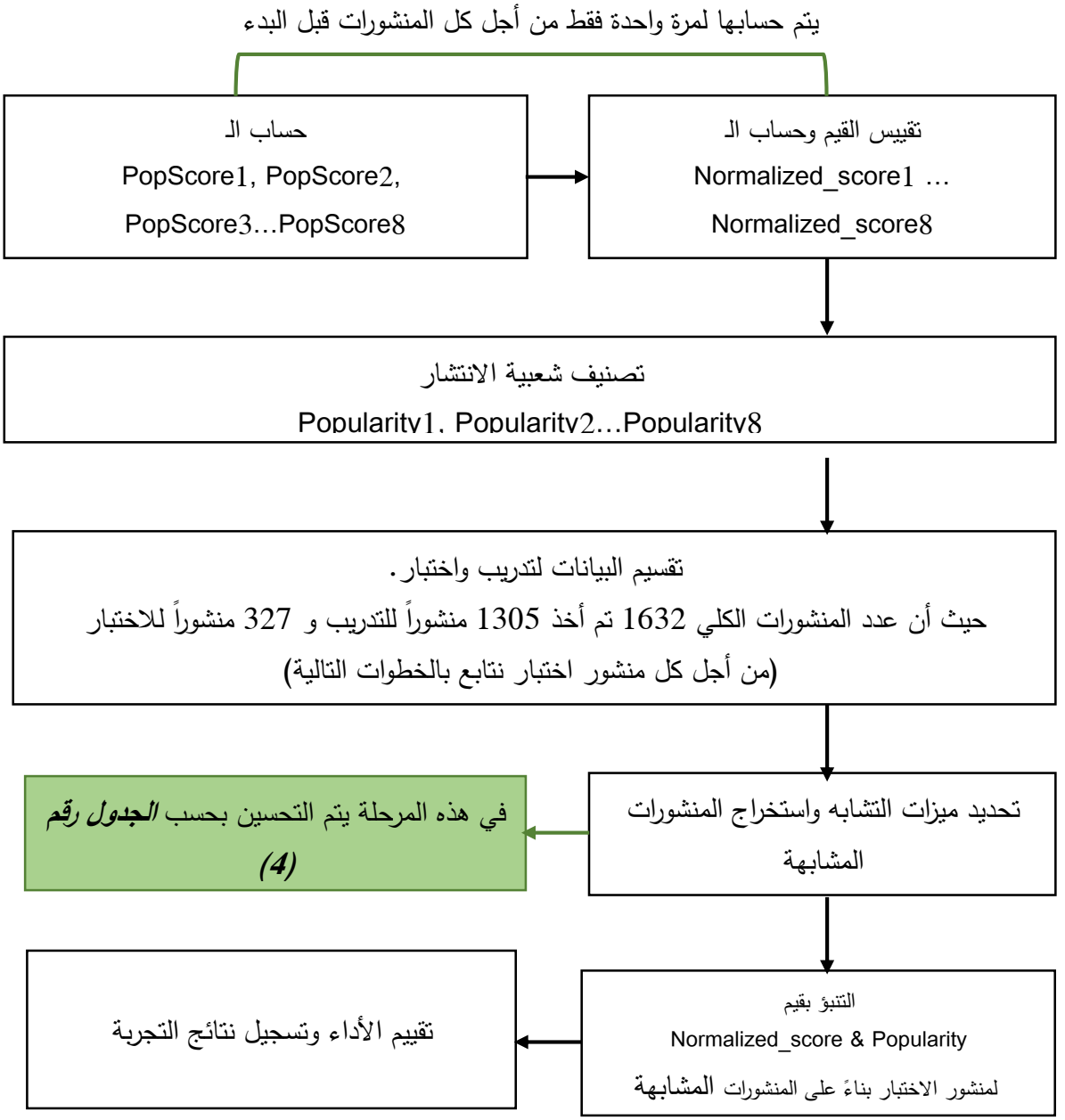
Normalized_score1, Normalized_score2,

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.

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

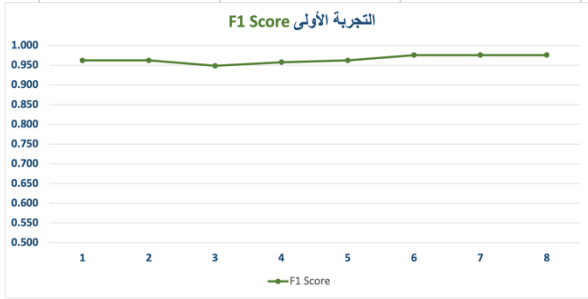
قد يخطر في البال تساؤلاً عن حالة عدم وجود منشورات مشابهة بنسبة 0.7 أو أكثر، في هذه الحالة يتم تضمين كافة المنشورات السابقة في حساب التنبؤ.

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

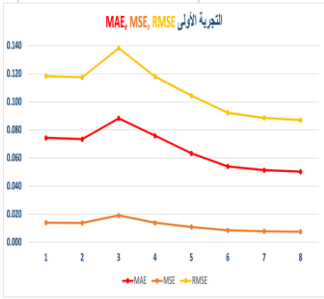
قيمته أعلى كانت الدقة أفضل، حيث أن نتائج التجربة الأولى موضحة في ***الجدول رقم (5):***

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.

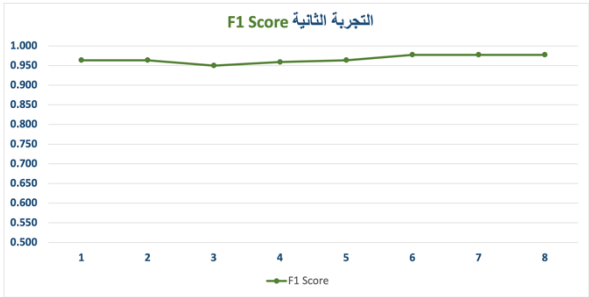
التجربة الثانية:

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)***.

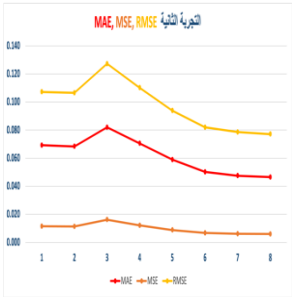
يظهر ***الجدول رقم (6)*** نتائج التجربة الثانية بعد التعديل على ميزات التشابه المدروسة في التجربة الأولى، كما أن الأشكال رقم ***(5)*** و ***(6)*** تظهر الشكل البياني لأداء المقياس.

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.