تحليل سلوك المستخدم واهتماماته من خلال معرض الصور الخاص به!

أسماء الطلاب:

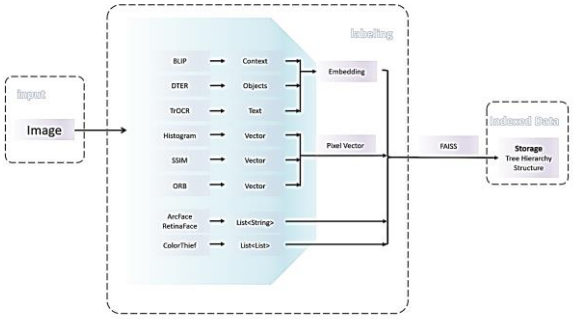
ليلى نصر
امير ماجد
سامي فرح
جان غزال

المقدمة:

نهدف في هذا المشروع لتحليل أنماط صور المستخدم المخزنة على جهازه الشخصي وفهمها، من خلال دراسة محتوى الصور ومحاولة تصنيفها لمجموعات تعكس اهتمامات المستخدم وهواياته والأماكن التي يرتادها عادةً والأطعمة التي يفضلها وأبرز الأحداث التي تثير اهتمامه، الخ..، ودراسة تأثير العوامل الخارجية على محتوى الصور وتحديد أكثر ما يشد اهتمام المستخدم وفترات الذروة التي يقوم المستخدم فيها بتخزين الصور
سيكون هدفنا فهم أنماط صور المستخدم من خلال دراسة محتوها context (لفهم الصور) والعناصر objects الموجودة فيها (من اشخاص، اثاث، حيوانات، أدوات الخ..)، وعدد هذه العناصر في الصور (الأمر الذي سيمكننا من استنتاج أهمها وأكثرها تكراراً بالمقارنة مع الأقل تكراراً والأقل اهمية)، والوجوه الموجودة في الصور وعددها (لنتمكن من استنتاج هوية صاحب هذه الصور والأشخاص الأقرب منه وإن كان اجتماعي ام انعزالي)، واستخراج النص من الصورة (حتى نستطيع التفريق بين الوثائق المخزنة كصور والصور التي يخزنها المستخدم من خلال كاميرة جهازه او من وسائل التواصل الاجتماعي)، وشرح عن خلفية الصورة) تكمن أهمية فهم الخلفية في فصل مكونات الخلفية background عن مكونات العناصر الأمامية foreground ويمكن بذلك استنباط معلومات جديدة مثل أكثر الأماكن التي يرتادها المستخدم(، أكثر الألوان الطاغية على الصورة (الميزة الذي قد نستطيع من خلالها فهم زمن نشاط المستخدم (نهاراً / ليلاً)، وعند ربطها مع محتوى الصورة نستطيع استنتاج معلومات جديدة)، والشعاع الممثل لجزئيات الصورة) pixel vector وهو الذي سنستطيع من خلاله تمثيل الصور فراغياً في نظام احداثي متعدد الأبعاد(high dimensional coordinate system
وذلك كله بغرض تصنيف الصور ضمن مجموعات، تساعد في تحليل سلوك المستخدم واهتماماته

المصدر: مجموعة البيانات VisionAI تم توليدها محلياً (يدوياً) من قبل المهندسة ليلى نصر

- تم جمع الصور من مواقع مفتوحة المصدر منشورة للعامة بدون حقوق ملكية
- تم استخدام منهجية مركبة من نماذج Vision-Language models و CNN models و Image preprocessing (Machine Learning) methods لعنونة مجموعة البيانات VisionAI كما هو موضح في الـ Table 1.
- تم استوحاء مجموعة البيانات هذه من احداث سنة 2024 في الجمهورية العربية السورية



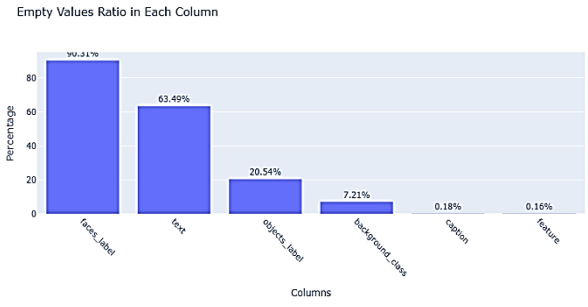
يوضح المنهجية المتبعة لعنونة الصور بغرض توليد الميزات المرتبطة بكل صورة :Table 1

شرح مجموعة البيانات:VisionAI

هذه المجموعة تتكون من:

1. مسار الصورة المخزنة على جهاز المستخدمpath: String
2. اسم الملف الحاوي للصورةfolder: String
3. تاريخ تخزين الصورة على جهاز المستخدمdata: datatime
4. نص يعبر عن محتوى الصورةcaption: long String
5. عدد الوجوه الموجودة بالصورةtotal_faces: int
6. أسماء الأشخاص التي تمثل الوجوه في الصورة :faces_label List<String>مدخلة عن طريق المستخدم
7. عدد العناصر الموجودة بالصورةtotal_objects: List<int>
8. عناوين العناصر الموجودة بالصورةobjects_label: Long String :
9. النص الموجود في الصورةtext: Long String
10. شرح عن خلفية الصورةbackgroung_class: Long String
11. الألوان الأكثر تأثيراً في الصورةcolor_palette: List<List<int>>
12. شعاع ممثل لميزات pixels الصورةfeature: array<int>
13. معرف المجموعة التي تنتمي لها الصورةcluater_id: int

معالجة مجموعة البيانات:



يوضح القيم المفقودة من الحقول في مجموعة البيانات :chart 1

1. معالجة القيم الفارغة :

نستنتج من المخطط chart 1. هنالك العديد من الأعمدة الفارغة مثل faces_labelحيث أنه العامود الذي يحوي أكثر القيم المفقودة وتصل لـ 90 بالمائة من مجمل عناصر مجموعة البيانات وهذا يدل على عدم وجود وجوه في هذه الصورة لذلك ملئنا القيم الفارغة بقيمة no_faces ولكننا قمنا بالتحقق

أولاً من وجود خطأ ما في عملية التعرف على الوجوه، وكان عامود total_facesمرجع لاستعادة
القيم الفارغة عن طريق الخطأ فيfaces_label ، فتحققنا في حال كانت قيمة الحقل total_faces اكبر أو تساوي واحد أي أن خوارزمية RetinaFace بالتقاط وجه لكن ArcFace لم تتعرف عليه ووفقاً للموقع الرسمي لـ DeepFaceفخوارزمية RetinaFace أفضل من ArcFace واكثر موثوقية، لذلك عندما وجودنا وجه أو أكثر فيtotal_faces ، قمنا باسناد قيمة الحقل المرافقة له في faces_label بأكثر وجه مكرر في العامود faces_label وهوrobin ، فكان أغلب النظر أن robin هو المالك لمجموعة البيانات هذه حيث تبلغ الصور التي تحتوي وجه robin حوالي الـ 92 بالمائة من مجمل الصور التي تحتوي وجه، وباقي الوجوه فتكاد ان تصل للـ 10 بالمائة من مجمل الصور التي تحتوي وجوه، واما عندما تكون قيمة الحقل صفر في total_facesفاسندناno_faces

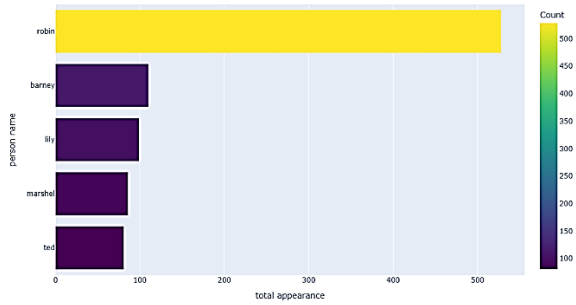


Chart 2: يوضح توزع اسماء وجوه الأشخاص في مجموعة البيانات

العامود text يحوي بنسبة 63 بالمائة من القيم المفقودة، وقد يكون السبب هو انه فقط 37 بالمائة من الصور هي صور لوثائق، وكسبيل للتأكد من صحة هذا الافتراض، قمنا باستعراض 10 صور من الصور التي تحتوي على نص لنجد أنها إعلانات لجامعات واقتباسات من كتب ووثائق وفواتير وقمنا كذلك باستعراض 20 صورة من الصور التي لا تحوي نص لنجد أنها خالية من الكلمات تماماً، لذلك قمنا بملئ العناصر الفارغة بـno_text

العامود objects_label يحوي فقط على 7 بالمائة من القيم المفقودة، يمكن استرجاع قيمه الفارغة ببساطة من شرح محتوى الصورة وشرح الخلفية، لكن بالنظر للحقول المرافقة للقيم الفارغة بالعامود total_objects نجد انها تحمل القيمة 0 أي انه لا يوجد أي عنصر، للتحقق من ذلك قمنا باستعراض 10 صور عشوائياً المرافقة للقيم الفارغة في objects_label لنجد أنها مناظر طبيعية ونصوص، لذلك قمنا بملئ القيم الفارغة بـno_objects

العامود caption و featureيحوي على سبعة قيم فارغة، لتعذر استخراج هذه الميزتان من Blip Vision Language ومن خوارزميات Machine Learningلعمليات معالجة الصور Image preprocessing هذا يعني بوضوح أن الصور المرافقة لهم غير مقروءة، أي لم يتم تحميلها من الشابكة بشكل صحيح اثناء عملية الـscraping ، لذلك قمنا بحذف هذه الحقول ببساطة

2. التأكد من صحة القيم:

في هذه المرحلة تم التأكد أن قيم الحقل total_faces صحيحة بالمقارنة مع faces_labelأي تم عنونة جميع الوجوه، والتأكد من أن قيم الحقل total_objectsصحيحة بالمقارنة مع objects_label أي تم عنونة جميع العناصر، والتحقق أن طول الأشعة الممثلة لجزئيات pixels الصورة) أي الحقل (feature جميعها متساوية، وكذلك بالنسبة لحقل color_palette أن لجميع الصور تم اسناد مصفوفة من 30 قيمة لونية (التي تمثل أبرز القيم اللونية في الصورة)، لنجد عدم وجود أي خلل

3. تغيير صيغة (نوع) القيم في الحقول لتصبح مقروءة بالنسبة لمكتبةpandas

وقمنا كذلك بتحويل قيم الحقل date لقيم مقروءة زمنية باستخدام to_datetimeوتحويل الحقول التي تحمل قيم من نوع مصفوفة لقيم مقروءة بشكل صحيح باستخدامast.literal_eval
الآن، يمكننا الانتقال للخطوة التالية

استكشاف مجموعة البيانات

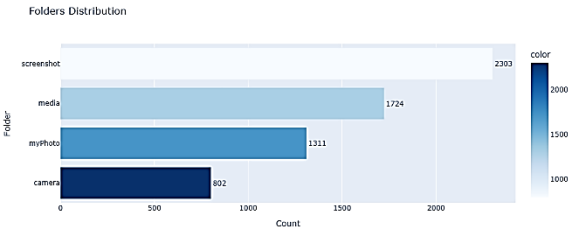


Chart 3: عدد الصور في كل مجلد

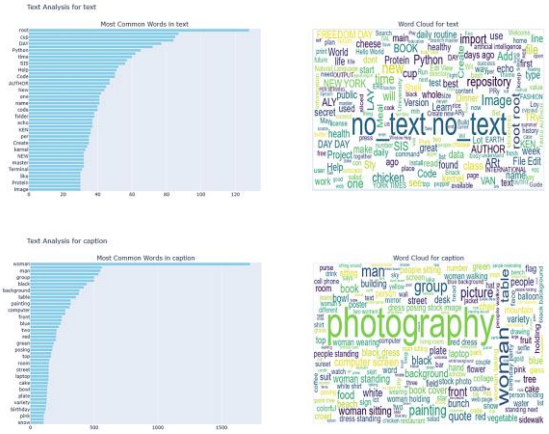
1- تحليل توزع الصور على الملفات:

قمنا بتجميع مجموعة البيانات وفقاً لحقل الـ folder وإحصاء مجمل عدد العناصر في كل من الملفات، ليظهر لدينا أربع ملفات ويمكن فهم بوضوح ثلاثة منها وهي(screenshot, media, camera) ، اما بالنسبة لملف myPhotoفهو مبهم المصدر سنتعمق في محتوى هذه الملفات في الخطوات القادمة، ولكن لاحظنا توزع صور المستخدم بشكل كثيف في ملف لقطات الشاشة والملف المسؤول عن تخزين الصور القادمة على الجهاز من مواقع التواصل الاجتماعيmedia ، وقلتها في ملف صور الكاميرا، وهذا يدل أن المستخدم لا يهتم بتوثيق يومياته خلال سنة 2024 واهتمامه بشكل كبير بتخزين ما يرده من محيطة (أصدقاء، عائلة، مواقع تواصل اجتماعي)، كما أنه يدل أنه يهتم بالعديد من المصادر خارج اطار الحياة الاجتماعية وذلك بسبب هيمنة ملفscreenshot

2- تحليل محتوى الصور والنصوص إن وردت فيهم

كما يظهر في المخطط .chart 4 قمنا بتجميع جميع النصوص والقيم الخاصة بشرح محتوى الصورة في نص واحد (أي متغير واحد)، وقمنا بعملية فلترة لهذا النص من خلال اقصاء كلمات التوقف stop words مثل in, with, you, me, at...)والمحارف المميزة (فواصل، إشارات استفهام، الخ..)، وقمنا بعملية اسناد كل كلمة وردت في النص بقيمة تعبر عن عدد تكرار هذه الكلمة واستعراض أكثر الكلمات تكراراً وكذلك استعراض النص من خلال word cloudلأخذ فكرة عامة عن محتوى الصور
لنستطيع تمييز خمس مكونات مهمة مكنتنا من التعمق بشخصية المالك:
تكرار كلمة women بشكل كبير جداً، مؤكداً ان كمية كبيرة من الصور هي لامرأة وبما أن الاسم الطاغي هورobin ، فهذا يدل بوضوح أن مالكة الصور هي فتاة تدعى robin

يمكننا استنتاج هوايات أو اهتمامات المستخدمة بسبب تكرار كل من : chess, drawing, dresses, Ken Follett author, science fiction books, art, AI topics
دالاً على أن robin تهتم بالرسم ولعب الشطرنج وقراءة كتب علمية وروايات أدبية تاريخية بوليسية لـ Ken Follet بسبب تكرار اسمه وأسماء رواياته بشكل كبير، كما أنها تهتم بالذكاء الصنعي والبرمجة لظهور العديد من مصطلحات البرمجة
يمكن استنتاج مهنة المستخدمة، بسبب تكرار كبير بنصائح خاصة بلغات برمجية ومعلومات ومواضيع تخص مجال الذكاء الصنعي، وتكرار كلمات تدل على أدوات الكترونية مثل (laptop – keyboard – screen...)
يمكن استنتاج الروتين الغذائي والأطعمة المفضلة لدى المستخدمة، بسبب تكرار كلمات مثل (salad, vegetables, healthy style, healthy routine tips...)
ويمكن تحديد القدرة الشرائية والوضع المادي، بسبب تكرار كلمة (bills, grocery shop list, fees...)فبتتبع الصور الخاصة بهذه الكلمات يمكن استخلاص هذه المعلومة
تم تجميع هذه الميزات لفئات باستخدام نموذج gpt o4 لتحليل ولاستخلاص معلومات منها بشكل يدوي



يوضح أكثر الكلمات تكراراً كمخطط حقول وكمخطط سحابي للكلمات في مجموعة البيانات :chart 4

3- تحليل خلفية الصور عند وجود وجوه في الصور:

بنفس آلية المهمة السابقة، قمنا بتجميع جميع النصوص الخاصة بشرح خلفية صور في ملف صور الكاميرا في نص واحد واستعراض اكثر الكلمات تكراراً واستعراضword cloud
من المخطط .chart 5 وبتحليل الكلمات نجد اربعة مناطق مميزة ترتادها المستخدمة بشكل كبير وهي:

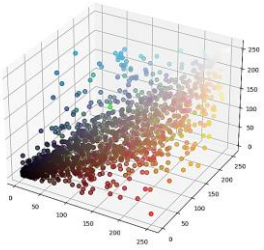
- المنزل
- المطاعم
- محلات الألبسة (النسبة الأكبر)
- متاجر السجائر
- وهنالك أماكن أخرى مثل: المكاتب، مكاتب الدراسةinternet booth

4- توزع القيم اللونية في ملف صور الكاميرا:

قمنا في هذه الخطوة باجتزاء القيم المرافقة لقيمة camera من الحقل foldersومن ثم اخذ الحقلcolor_palette ، وبعدها بتطبيق قاعدة استخلاص كمية الإضاءة luminance formula من القيم اللونية RGB كما هو موضح في المعادلة (1) لكل لون في مصفوفة الألوان المرفقة بكل صورة،

$$(1)\quad Luminance = 0.2126 * R + 0.7152 * G + 0.0722 * B$$

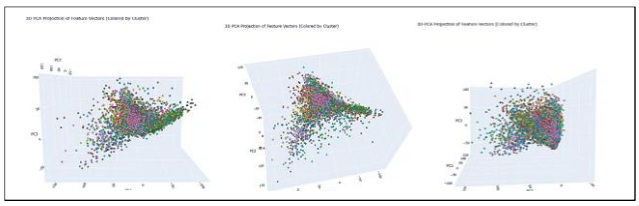
استطعنا الفهم من المخطط ثلاثي الابعاد hologram 1.متى تم التقاط هذه الصورة (فترة الصباح ام الليل)، لنجد أن معظم الصور تم التقاطها بظروف إضاءة منخفضة، لنستنتج أن المستخدمة robin تنشط مساءاً، فتقوم بممارسة نشاطاتها وبارتياد اماكنها المفضلة (التي وجدناها في الخطوات السابقة) في وقت المساء



تمثيل القيم اللونية للصور في مجموعة البيانات :Holoaram 1

5- تحليل توزع الأشعة الممثلة للصور:pixels vector

هذه الخطوة هي مجرد خطوة أولية لعملية تدريب نماذج التعلم التلقائي على مجموعة البيانات، فكخطوة مبدئية كان علينا استعراض الصور على شكل اشعة في نظام احداثيات متعدد الأبعاد high dimensional coordinate system، لنجد أن مجموعة البيانات من النوع الذي يمتاز بكثافة عالية high densityوالميزات تتوزع فيه بشكل هرمي hierarchal data وبعد أن استعرضنا اشعة الصور pixels vector بالنسبة لحقل معرف المجموعة cluster_idالمنتمية له، نجد أن مجموعات الصور متداخلة بشكل كبير وتحوي على العديد من القيم المتطرفة outliers التي تحتاج لعملية فصل

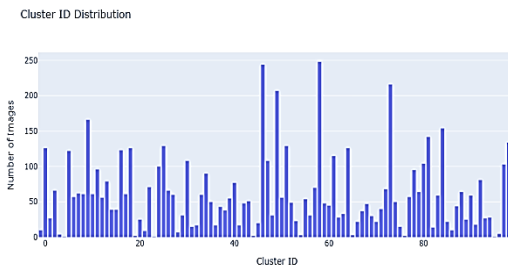


تمثيل ثلاثي الابعاد لأشعة الصور تبعاً لتوزعها ضمن العناقيد المبدئية :Hologram 2

6- تحليل المجموعات (العناقيد) المصنفة للصور:

- دراسة توزع مجمل عدد الصور في كل مجموعةcluater_id

قمنا في هذه المرحلة بتجميع مجموعة البيانات وفقاً للـ cluster_id واحصاء عدد العناصر في كل مجموعة، وفقا للمخطط .chart 6 والاحصائيات describeلاحظنا وجود 101 مجموعة تختلف بشكل شاسع بما يخص مجمل عدد الصور في كل منها، حيث أن توزعها standard deviation يبلغ 52 صورة ووسطياً تحتوي المجموعات على 60 صورة بينما 75 بالمائة منهم يحوي بما يقارب الـ 80 صورة، أما هنالك مجموعات تحوي ما يقارب الـ



مجمل عدد الصور في كل عنقود من مجموعة البيانات :Chart 6

250 صورة، وهذا دليل على تشتت مجموعات البيانات بشكل كبير بما يخص مجمل عدد الصور، ونقترح هنا إمكانية دمج بعضها ببعض، وستتضح هذه الامكانية في الخطوات التالية

- **توليد ميزة جديدة للدلالة على كثافة المجموعة:**

بعد ما قمنا به بالخطوة السابقة من دراسة لتوزع الصور على المجموعات، قمنا بفرز المجموعات لثلاث فئات، حيث وسطياً نجد ان المجموعات تحوي من 60 – 150 صورة فسنعتبرها مجموعات ذات كثافة معتدلة moderate density فإن قل مجموع الصور عن 60 سنعتبر المجموعة ذات كثافة منخفضة low density وإن زاد المجموع عن 150 سنعتبر المجموعة ذات high density

- **علاقة المجموعات cluster_id بالوجوه total_faces**

تتوزع عدد الوجوه على المجال [0 – 22] على الرغم من قلة عدد الصور التي تحتوي على عدد وجوه اكثر من 3، إلا أن ثقتنا بنموذج RetinaFace كبير، وإن تنبئ بوجود 100 وجه لن نشك به، لذلك سنكمل بدون القيام بأي عملية drop

لنجد وفقاً للمخطط chart 7. انفصال شاسع للمجموعات cluater_id بين صور تحتوي على ما يزيد عن 200 وجه ومجموعات خالية تماماً من الوجوه، فهنالك فقط ثلاث cluater_ids تحوي على الكثير من الوجوه

- **توليد ميزة جديدة للدلالة على عدد الوجوه في المجموعات**

على الرغم من إمكانية استنتاج فئتان فقط لفصل المجموعات لـ (no_faces, has_faces) إلا أن الإحصائيات توجبت وجود أربع فئات إذ أن العدد الأكبر من الصور يحوي فقط على وجه واحد، لذلك يجب مراعات المجموعات التي قد تحوي على عدد قليل من الصور التي تحوي وجوه وبالتالي لا يوجد مجال لتصنيفها مثل ما قد نصنف المجموعات التي تحتوي على ما يتجاوز ال 200 وجه

لذلك تم اعتبار الحد الأدنى هو وجود وجه – 50 وجه واعتبار حد وسطي هو وجود اكثر من 50 وجه ولكن اقل من 200 وحد اعلى عند تجاوز ال 200 وجه، ووسم المجموعة على انها no_faces في حال عدم وجود وجوه اطلاقاً

- **دراسة العلاقة بين المجموعات cluster_id والعناصر الموجودة فيها objects**

لإنجاز هذه الخطوة سنقوم بجمع جميع العناصر الخاصة بالمجموعة الواحدة cluater_id بنص ونقوم باستخراج اكثر العناصر objects تكراراً، وجب التنويه: أن عدد ال objects الكلي يتجاوز الـ 80 عنصر، وبالتالي استنتاج العناصر اكثر تأثيراً بكل مجموعة لن يساعدنا بفهم فحوى المجموعة، فقمنا بتجميع العناصر المتشابهة بإطار شامل لهم، وبذلك نعمم على المجموعة نوع معين من العناصر

باستخدام خوارزمية تصنيف gpt o4 للعناصر بالاعتماد على كمية تكرارها وإلى ما ترمزه عندما توضع في سياق، تولد لدينا ما يلي من الفئات الخاصة بهم:

- حيوانات
- أطعمة
- عناصر توجد في داخل الأبنية Indoor
- عناصر توجد بالطبيعة outdoor
- أجهزة الكترونية
- أدوات رياضية
- أدوات شخصية
- اشخاص
- كتب
- غير معرف

ومن ثم تعميم فئات العناصر هذه على المجموعات، فاستطعنا حصر ما يزيد عن 80 عنصر مختلف فقط بما يقل عن 10 فئات

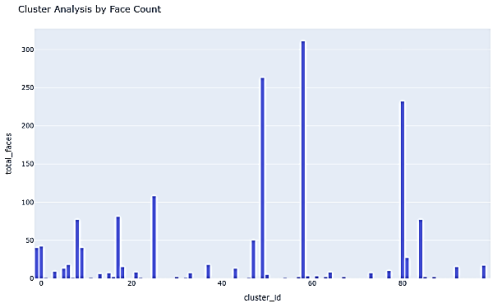


Chart 7: عدد الوجوه في كل عنقود من مجموعة البيانات

- **بناء ميزة جديدة للدلالة على فئة العناصر الأعم في كل مجموعة**

على الرغم من ان توحيد 80 عنصر في ما يقارب 10 فئات، إلا أن معظم المجموعات ما زالت تعاني من تعدد الفئات، لذلك توجب وسمها بفئة العناصر الأعم أي التي تضم اكبر عدد صور، ولحسن الحظ لم تتساوى أي فئتين لأي مجموعة كانت، فتمت عملية تعميم الفئة بنجاح

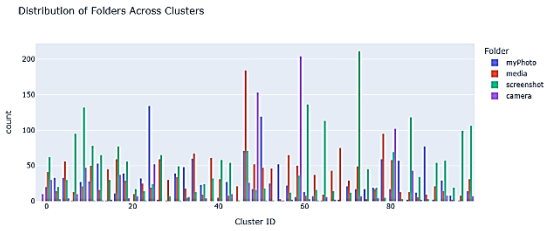


Chart 8: عدد الصور في كل عنقود مجمعة وفقاً للمجلد التي تنتمي له

- **وبنفس الطريقة تم دراسة وبناء ما يلي:**
- توزع الملفات على المجموعات، كما هو موضح في chart 8.
- بناء ميزة جديدة للدلالة على المجلد المهمين في المجموعة
- توزع خلفيات الصور على المجموعات، وجب التنويه هنا: أن كان التحدي أصعب من أي شيء مر معنا من قبل، لأنه تم التعامل مع ما يقارب 2000 ألفين خلفية مختلفة☹
- توليد ميزة جديدة للدلالة على فئة الخلفيات الأعم في كل مجموعة
- توزع محتوى الصورة على المجموعات، على الرغم من اخلاف المحتوى لكل صورة، انما هنالك ترابط كبير بالمحتوى وبالمفردات بين الصور المتشابهة وكان من السهل اكتشافها وعزلها
- توزع النصوص بالصورة ان وجد في مجموعات البيانات، وكان مفيد جداً بعملية عزل الوثائق، وصور الشاشة لأسطر وخوارزميات برمجية ومواقع على الشابكة وكذلك نصائح واقتباسات، الخ..
- توليد ميزة جديدة لفئة الكلمات الأكثر تأثيراً بالمجموعة

في نهاية هذه المرحلة: قمنا ببناء مجموعة بيانات cluster_data تحتوي على معرفات المجموعات id (ستفيدنا بعملية الدمج مع مجموعة البيانات الاساسية)، ومجمل عدد الصور في كل مجموعةtotal_image ، وكثافة المجموعة cluster_density وفئة العناصر الأعم في المجموعة commom_object_label، والملف الأعم للصور في المجموعةdominant_folder ، واعم خلفية للصور common_background_class واعم محتوى للصورcommon_context_theme

7- تقليل عدد المجموعات المصنفة للصور:

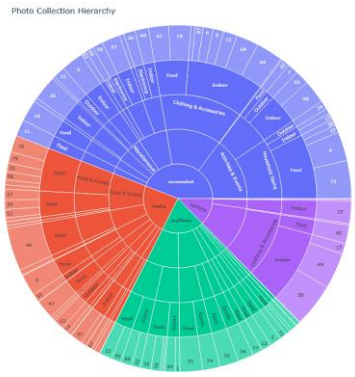


Chart 9: يوضح تصنيف العناقيد وفقاً لنوع المجلد وأبرز ما ورد فيه من عناصر وخلفيات ونوع محتوى الصور فيه

بعد ما تم إنجازه من معاينة دقيقة وعميقة للمجموعات، كانت المفاجئة أنه حقاً يمكن تخفيف عدد المجموعات بسبب تشابه معظمها، ونجحنا بتخفيض عدد المجموعات من 101 لما يقارب 20 مجموعة، وفقاً لدراسة يدوية لتشابه هذه المجموعة من خلال مخططsunburst

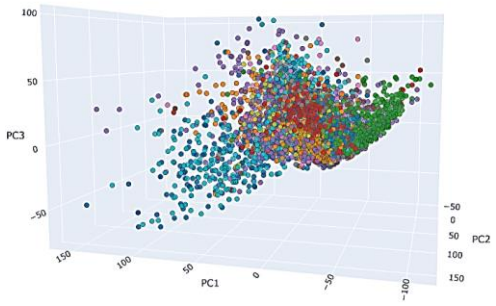
على الرغم من مظهر sunburst المخيف الموضح في chart 9. إلا أن باستخدام مكتبة plotly والخدمات التفاعلية التي تقدمها، قمنا باستكشافه يدوياً واستخلاص المجموعات المتشابهة منه وفقاً للجدول الآتي:

Duplicate Cluster Patterns (16 Total)

Folder	Clusters	Categories (context_theme → background_class → object_label)
Screenshot	84, 64, 61, 13, 8, 6, 87, 20	Indoor → Clothing & Accessories → Clothing & Accessories
	40, 26	Indoor → Household Items → Technology & Office
	90, 97	Indoor → Electronics & Machines → Technology & Office
	94, 15, 73	Indoor → Household Items → Activities & Events
	9, 38	Food → Household Items → Activities & Events
	42, 37	Food → Building & Structure → Activities & Events
Media	78, 79, 35, 68, 62, 99	Food → Food & Drink → Food & Drink
	23, 45, 52, 39, 27	Food → Nature & Plants → Food & Drink
	10, 66	Outdoor → Building & Structure → Miscellaneous
	48, 57	Indoor → Household Items → Household Items
	33, 55, 75, 86	Food → Household Items → Activities & Events
	1, 83	Indoor → Household Items → Activities & Events
	88, 50, 43	Indoor → Clothing & Accessories → Activities & Events
Camera	80, 49	Food → Clothing & Accessories → Activities & Events
	17, 58	Indoor → Clothing & Accessories → Activities & Events

والآن السؤال الذي راودنا باستمرار من بداية هذه المرحلة (التي تتكون من 13 خطوة) هل عملية اختصار المجموعات هي عملية مفيدة ام مدمرة لهيكلية مجموعة البيانات وخطأ فادح من البداية، نجد أن هيكلية البيانات مصانة من خلال استعراض توزع اشعة الصور على المجموعات

وليس هذا فقط، إنما عزل للقيم المتطرفة أيضاً 😊 كما هو موضح في التمثيل hologram 3.



Hologram 3: تمثيل ثلاثي الابعاد لأشعة الصور في مجموعة البيانات وفقاً لتوزعها ضمن العناقيد الجديدة

8- دراسة السلسلة الزمنية لتوزع الصور شهرياً ويومياً على مدار سنة 2024

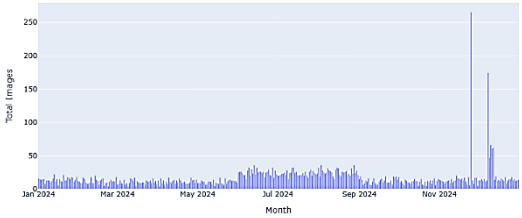


Chart 11: توزع عدد الصور وفقاً لأيام سنة 2024

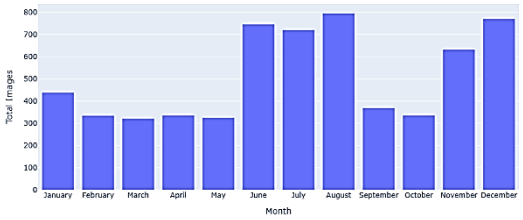


Chart 10: يوضح توزع عدد الصور تبعاً لأشهر سنة 2024

لدراسة توزع عدد الصور على مدار السنة شهرياً، قمنا ببناء ميزة جديدة وهي الشهر من خلال استخلاصها من الحقلdate ، واستعراض عدد الصور بكل شهر وفقاً للمخطط.chart 10 ، لنجد ارتفاع كبير فيهم خلال اشهر الصيف، للدلالة على زيادة نشاط المستخدم خلال هذه الفترة بما يعادل ضعفي نشاطه بباقي الأشهر

إلا أن كل من شهر November وشهر December يبدون ارتفاع كبير وملحوظ مقارنة بأشهر الشتاء والخريف
بتحليل عدد الصور المخزنة خلال أيام السنة وباستعراضها وفقاً للمخطط chart 11.، نجد أن زيادة عدد الصور في الشهر 11 و12 بسبب يومين وهما Nov 25 و Dec 8

فما الذي حدث في Dec 8 عزيزي السوري الحر هل بسبب زيادة شرب المرطباط مثلا؟؟؟

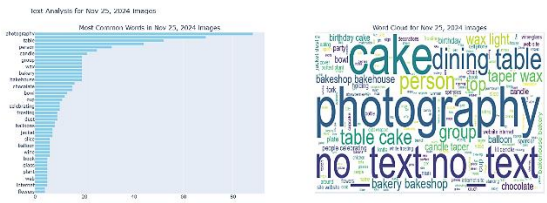


Chart 12: اكثر الكلمات تكراراً في الخامس والعشرين من شهر نوفيمبر

Nov 25: بتحليل العناصر objects الموجودة بالصور ودراسة خلفيتهم وبتحليلي النصوص ان وجدت بالصور ومحتواهمcontext ، فمن خلال جمع ما سبق من ميزات تخص صور Nov 25 في نص واحد وتحليل مفرداته ومن المخطط chart 12. ، نجد تكرار كبير في الكلمات التالية:

Cake, candles, birthday cake, group of people, celebrating, balloons, slice…

والكثير من الكلمات للدلالة على عيد مولد مالكة مجموعة الصور، فهذا يبرر ارتفاع كمية الصور المخزنة في هذا اليوم تحديداً، وذلك بسبب وفد صور من مختلف الوسائط إلى المجلدات الموجودة على جهاز المستخدمة احتفالاً بها

Dec 8: بنفس ما تم إنجازه في الخطوة السابقة وعرض المخطط chart 13. بغرض تحليل الصور، تم العثور على أكثر المفردات تكراراً وهي:

Crowd, flags, military, firework, uniform, national, Freedom Day, Syrian flag, national freedom

وبما ان مجموعة البيانات جمعت بعام 2024 بما يتوافق مع مجريات احداث هذه السنة في الجمهورية العربية السورية، فسبب ارتفاع عدد الصور بهذا الشكل على الأيام الممتدة بين Dec 8 – 12 هو يوم تحرير سوريا من النظام الفاسد

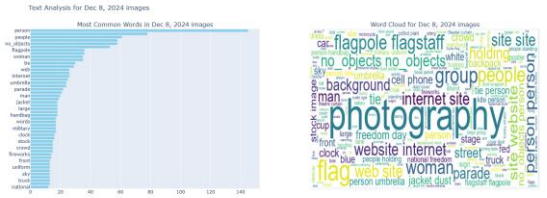
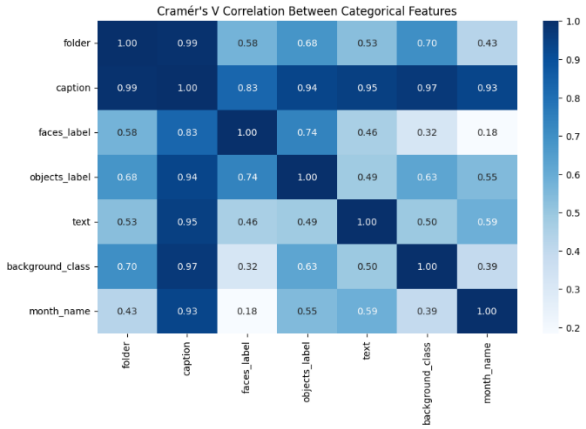


Chart 13: أكثر الكلمات تكراراً في الثامن من شهر ديسمبر عام 2024

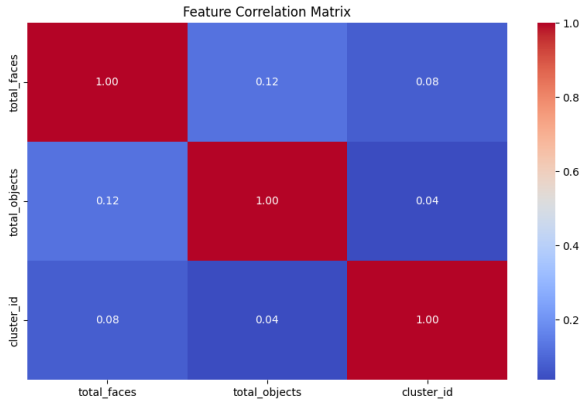
9- الارتباط الخطي للميزات بين بعضها

لتحليل الارتباط الخطي، قمنا بفصل الميزات الفئوية عن الميزات الرقمية، وقمنا بحذف الميزات التي تحمل قيم مميزة لكل صورة وهم feature, color_palette, pathوذلك لأنها لن تساعد في كشف أنماط الارتباط وستحقق ارتباط عالي مع أي ميزة

بتطبيق خوارزمية كريميرز Cremers V وجدنا ارتباط خطي من متوسط 0.3 – 0.5 لكبير 0.6 – 0.9 بين الميزات الفؤية



وارتباط ضعيف لا يتجاوز 0.1 بين الميزات الرقمية، عند تطبيق خوارزمية Pearson



اعداد مجموعة البيانات لعملية التدريب

1- توليد ميزات جديدة

كان من الضروري توليد ميزات جديدة بسبب عدم قدرة تدريب نماذج التعلم العميق على بعض الميزات التي تحمل قيم من نوع مغاير لنص او رقم او فئات (بما معناه تحمل قيم من نوع مصفوفة مركبة) مثل color_palette, total_objects

وكهدف رئيسي من تحليل المصفوفة اللونية لكل صورة، هو فهم التوزع اللوني فيها وكمية الإضاءة brightness فيها

وشدة القيم intensity اللونية ومقدار اختلاف variance القيم اللونية فيها، كما هو موضح في كل من المعادلات(1, 2, 3)

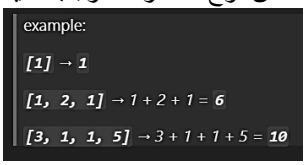
$$(2)\ \boldsymbol{Variance} = \boldsymbol{Standard\ Daviation}(\boldsymbol{R}+\boldsymbol{G}+\boldsymbol{B})$$

$$(3)\ \boldsymbol{Intinsity} = (\boldsymbol{R}+\boldsymbol{G}+\boldsymbol{B})/3$$

لذلك قمنا باستخراج هذه المعلومات، لما لهم من دور مهم في عملية فهم أنماط توزع وتشابه الصور، وحذف حقل المصفوفات اللونية

لاحظنا من دراستنا السابقة وجود العديد من المحتويات التي تحوي مفردات مثل (boy – women – man – girl – group – crowd – person – people – someone …)والعنصر object (person) الذي يدل على وجود شخص ما في الصورة، وبالتالي نحن بحاجة لالتقاط هذه المعلومة لما لها تأثير على أنماط تصنيف الصور، لذلك انشأنا حقل جديد has_person وقمنا بإسناد قيم بوليانية له

بسبب صعوبة عملية الـ embedding لقيمة من نوع مصفوفة، توجب علينا تحويل الحقل total_objects لحقل قيم رقمية من خلال جميع عدد العناصر في المصفوفة، وفقاً للمثال الموضح :



2- دراسة الارتباط الخطي بين الميزات والمجموعات

بتطبيق خوارزمية Cremer's V على القيم الفئوية وخوارزمية ANOVA Eta-squered وبعد تحول قيم حقل cluster_group وحذف الحقول التي تحوي قيم مميزة على طول مجموعة البيانات (caption, text, objects_label, background_class, feature) لنجد ما يلي:

	feature	correlation_type	value	interpretation
9	dominant_folder	Cramer's V	0.816052	categorical
10	common_background_class	Cramer's V	0.728175	categorical
11	common_context_theme	Cramer's V	0.728134	categorical
7	faces_categories	Cramer's V	0.712885	categorical
6	cluster_density	Cramer's V	0.634028	categorical
0	folder	Cramer's V	0.577764	categorical
8	common_object_label	Cramer's V	0.549679	categorical
1	faces_label	Cramer's V	0.199188	categorical
3	cluster_id	ANOVA Eta-squared	0.034625	numerical
15	has_person	ANOVA Eta-squared	0.013527	numerical
2	total_faces	ANOVA Eta-squared	0.010506	numerical
14	color_brightness	ANOVA Eta-squared	0.003689	numerical
16	total_object_int	ANOVA Eta-squared	0.003503	numerical
12	color_intensity	ANOVA Eta-squared	0.003330	numerical
4	month	ANOVA Eta-squared	0.002050	numerical
13	color_variance	ANOVA Eta-squared	0.001704	numerical
5	day	ANOVA Eta-squared	0.001096	numerical

ارتباط خطي كبير يصل لـ 0.6 وسطياً بين cluster_group والقيم الفئوية وارتباط ضعيف جداً يكاد لا يزيد عن 0.001، لذلك سنستبعد القيم الرقمية

3- تحضير مجموعة البيانات لعملية التدريب

- قمنا بإضافة 1500 صورة معنونة لم يتم فحصها او تحليلها سابقاً لاختبار نتائج التدريب عليها
- قمنا بتحويل القيم الفئوية لصيغة مقروءة من قبل نماذج التعلم التلقائي باستخدام OneHotEncoder
- قمنا بتضمين حقل الـ feature كخطوة نهائية على مجموعة التدريب ليكون embedding للصور
- تم دمج خرج preprocesser OneHotEncoder و Feature Embedding بطريقة stacking
- فصل عشوائي مع stratified للحفاظ على شكل المجموعات cluster_group لمجموعة البيانات (تحوي ما يقارب 6500 صورة) التي تم تحليلها سابقاً لمجموعة train ومجموعة validation
- حيث تتألف مجموعة التدريب مما يقارب 5000 عنصر
- ومجموعة التحقق مما يقارب 1500 عنصر

تدريب مجموعة البيانات

1- التقنيات والتحسينات المستخدمة (Optimization Techniques Applied)

تم تطبيق مجموعة من الإجراءات المتقدمة لتحسين جودة البيانات وأداء النماذج:

- **معالجة القيم المفقودة:(Missing Values Handling)**
تم استخدام خوارزمية **SimpleImputer** لتعويض القيم المفقودة إما بالمتوسط أو النمط.(mean/mode)
- **تقييس البيانات:(Feature Scaling)**
استخدام **StandardScaler** لتقييس المتغيرات العددية وتحسين استقرار النماذج التي تتأثر بحجم السمات.
- **هندسة الميزات:(Feature Engineering)**
اشتملت على استخراج ميزات نصية باستخدام **TF-IDF**، بالإضافة إلى تحليل النصوص واستخدام تقنيات استخراج السمات من الأعمدة النصية والتاريخية.
- **ضبط المعاملات:(Hyperparameter Tuning)**
تم استخدام أدوات مثل **GridSearchCV** أو **Optuna** لتحسين أداء النماذج عن طريق اختيار أفضل تركيبات المعاملات.

2- إعداد وتدريب مجموعة البيانات

تم تقسيم البيانات إلى ثلاث مجموعات رئيسية:

- مجموعة التدريب (X_train_final, y_train_final): لاستخدامها في تدريب النماذج.
- مجموعة التحقق/التقييم (Validation): استخدمت خلال عملية ضبط المعاملات باستخدام Optuna.
- مجموعة الاختبار النهائية (X_test_final, y_test_final): لتقييم الأداء العام للنماذج بعد التدريب وضبط المعاملات.

قبل بدء التدريب، تم تطبيق بعض الخطوات التحضيرية المهمة:

- تحجيم الميزات (StandardScaler) لتوحيد مقياس البيانات، وهو أمر ضروري خاصةً للنماذج التي تعتمد على حساب المسافات مثل KNN.
- ترميز الفئات (LabelEncoder) لتحويل التصنيفات النصية إلى أرقام مناسبة للنماذج

النموذج	الدقة (Accuracy)	F1 Score (Weighted)	ملاحظات
Logistic Regression	0.5469	0.5179	أداء ضعيف جدًا
K-Nearest Neighbors (KNN)	0.3527	0.3414	أسوأ نموذج
Random Forest	0.9339	0.9257	أداء ممتاز
XGBoost	0.9535	0.9499	دقة ممتازة مع تحذير بسيط overfitting
LightGBM	0.9649	0.9626	الأفضل أداءً حاليًا

Table 3: مقارنة نتائج نماذج التعلم بعد التدريب على مجموعة البيانات من حيث الدقة

3- اختيار النماذج وأسباب ذلك

تم اختيار مجموعة متنوعة من الخوارزميات لتعكس تنوعًا في الأساليب والاستراتيجيات:

- **K-Nearest Neighbors (KNN):** نموذج بسيط يعتمد على مبدأ القرب الجغرافي (المسافات)، مناسب للبيانات الصغيرة نسبيًا.

- **XGBoost:** نموذج متقدم قائم على أشجار القرار ويعتمد على تقنية التعزيز التدريجي(Gradient Boosting) ، فعال جدًا في التعامل مع التصنيفات المعقدة والمتعددة.
- **LightGBM:** مشابه لـ XGBoost لكنه يبرز بسرعته العالية وكفاءته في التعامل مع مجموعات بيانات كبيرة ومتعددة الفئات.
- Random Forest: نموذج يعتمد على تجميع أشجار القرار (Ensemble)، يوفر أداء مستقرًا مع قابلية جيدة للتعامل مع البيانات المتنوعة دون الحاجة لضبط معقد.
- **Logistic Regression:** نموذج خطي بسيط يُستخدم كأساس للمقارنة، لكنه أقل فعالية مع البيانات غير الخطية أو غير المتوازنة.

بالتالي، تم اختيار النماذج لتمثيل استراتيجيات مختلفة مثل:

- النماذج المعتمدة على الجوار(KNN)
- نماذج التعزيز باستخدام الأشجار(XGBoost) ، (LightGBM
- نماذج التجميع(Random Forest)
- النماذج الخطية(Logistic Regression)

ويمكن رؤية أداء وسرعة تنفيذ ومدى تعقيد كل من النماذج السابقة في الجدول Table 6.

النموذج	الدقة (Accuracy)	F1 Score (Weighted)	CV Accuracy	زمن التدريب
KNN (Optuna + GridSearch)	> 0.95	> 0.94	≈ 0.93	متوسط
XGBoost	0.9535	0.9499	فشل بسبب فئة نادرة جدًا	ثانية 127.7
LightGBM	0.9649	0.9626	-	سريع جدًا (6.28 ث)
Random Forest	0.9339	0.9257	مستقر	متوسط
Logistic Regression	0.5469	0.5179	مستقر	سريع

Table 6: مقارنة بين نماذج التعلم من حيث الأداء والسرعة التنفيذ

4- معايير التقييم وأسباب اختيارها

تم الاعتماد على مقاييس تقييم تعطي صورة شاملة للأداء، خصوصًا في ظل التوزيع غير المتوازن للبيانات:

- **دقة التصنيف :(Accuracy)** مقياس عام لكن قد يكون مضللاً في حال وجود توازن غير متساوٍ بين الفئات.
- **مؤشر :F1 Score (Weighted)** المؤشر الرئيسي، لأنه يجمع بين الدقة والاسترجاع مع مراعاة حجم كل فئة، مما يعكس أداء النموذج بشكل أكثر عدالة.
- **التحقق المتقاطع :(Cross-Validation)** لتقليل التحيز الناتج عن تقسيم واحد للبيانات وضمان استقرار النتائج.

تم استخدام Optuna لضبط المعاملات عن طريق تعظيم F1 Score ، ثم تقييم النماذج باستخدام هذه المعاملات على مجموعة الاختبار، كما نجد نتيجة تطبيق معايير التقييم على نماذج التعلم بعد تدريبها في الجدول .Table 3

ملاحظات:

- نموذج XGBoost واجه مشكلة أثناء التحقق المتقاطع بسبب وجود فئة نادرة جدًا (ظهرت مرة واحدة فقط)، ما أدى إلى فشل إحدى تجارب.CV

	التقييم
أفضل نموذج	**LightGBM** تفوّق على الجميع بدقة **96.49%** و**0.9626** ≈ F1، مع تدريب سريع (6 ثواني فقط) وأداء ممتاز على جميع الكلاسات. **XGBoost** بنتائج قريبة جدًا، لكن يحتاج مزيدًا من الضبط، ودلائل بسيطة على overfitting. **Random Forest** كان ممتازًا جدًا ويُعتبر خيارًا موثوقًا بدون تعقيد التخصيص العالي.
نماذج ضعيفة	**Logistic Regression** و**KNN** لم يتعاملا جيدًا مع البيانات متعددة الكلاسات وغير المتوازنة.

Table 4: المزيد من المعلومات عن كل نموذج بناءاً على ما تم إنجازه في مرحلة التقييم

- LightGBM تفوق من حيث الدقة، سرعة التدريب، ومؤشر F1 Score، وهو النموذج الأفضل بين جميع النماذج.
- KNN قدم أداءً جيدًا بعد التوليف، لكنه حساس لتحجيم البيانات ويحتاج إلى وقت أكبر أثناء التنبؤ.
- Random Forest حقق أداء قويًا ومستقرًا بدون الحاجة إلى ضبط معقد، مما يجعله خيارًا موثوقًا.
- Logistic Regression كان الأداء ضعيفًا نسبيًا، خاصة مع تعدد الفئات وعدم توازن البيانات.

المزيد من التفاصيل عن كل نموذج من حيث التقييم تم ادراجها في الجدول .Table 4

5- تحليل النموذج الأفضل LightGBM :

الأداء:

- الدقة: 96.49%
- F1 Score: 96.26%
- زمن التدريب: 6.28 ثانية فقط
- تقرير التصنيف يظهر أداء ممتازًا حتى مع الفئات الصغيرة الحجم

الأسباب التي تفسر تفوق LightGBM:

- فعالية عالية مع البيانات ذات الأبعاد الكبيرة والمتعددة.
- قدرة قوية على التعامل مع بيانات غير متوازنة.
- ميزة early stopping التي تمنع الإفراط في التعلّم (overfitting).
- تحسينات خوارزمية التدرج (Gradient-based) التي تعزز الدقة وتسريع التدريب.

النتيجة

بعد تحليل ميزات مجموعة البيانات، تبين أن المجموعة من النوع عالي الكثافة High dimensional وتتخذ شكل هرمي، واستخلصنا اهتمامات المستخدم، وفترات نشاطه ومن الممكن تنبئ بكمية البيانات ونوعها على فترة أشهر إلى الأمام، واستطعنا تقليل عدد المجموعات التي فرزت على أساسها مجموعة البيانات من خلال تطبيق عمليات هندسة الميزات بغرض إيجاد التشابه بينها، ونجد تفوق العديد من نماذج التعلم العميق في عملية تصنيف العديد من الأنماط الخاصة بسلوك واهتمامات المستخدم، أبرزها LightGBM حيث حقق أفضل نتائج تقييم ووصلت قيمة f1-score إلى ما يقارب 96% على مجموعة البيانات VisionAI

دراسة مرجعية

قدم البحث [1] يطرح هذا البحث فكرة قدرة نموذج K mean بتجميع القيم اللونية pixels إلى مجموعات فيتم بالتالي تقليل عدد الأطياف اللونية في الصورة تبعاً لرغبة المستخدم بعدد المجموعات K كما أن هنالك حد لن يحدث بعده الكثير من الفرق عند زيادة قيمة K، ما دفعنا لوضع هذه الدراسة ضمن الدراسات المشابهة هو أنها كانت من أقدم الخوارزميات والخطى نحو عملية تجميع الصور، فهي تقترح انه يمكن فهم توزع القيم اللونية وامتدادها واستخراج الحواف ومن ثم الزوايا من الصورة وبعض ال features البدائية بعملية معالجة الصور، كل ذلك بغرض بناء شعاع يعبر عن الصورة وتمثيلها فراغياً من ثم إيجاد مشابهاتها أو تصنيف الصور ضمن مجموعات

اهتم البحث [2] بتحسين التعلم الموحد (FL) لتحليل سلوك المستخدم باستخدام آلية تخزين مؤقت ذات مستويين(FedTCM) وهو نموذج موزَّع يحافظ على خصوصية البيانات، مما يجعله مناسبًا لتحليل سلوك المستخدم عبر مصادر متعددة، لكن اختلاف توزيعات البيانات بين المصادر (Non-IID) يؤدي إلى تحيز في التدريب، مما يؤثر على دقة النموذج وسرعة تقاربه.
حيث يتم تجميع المستخدمين في المنظومة المراد دراستها بناءً على تشابه توزيعات بياناتهم لتقليل تأثير Non-IIDداخل كل مجموعة، من ثم يستخدم الاتصال غير المتزامن بين الخوادم والعملاء للتغلب على اختلاف سرعات لحساب بين الأجهزة، ويمر النموذج بمستويين، **المستوى الأول** :تخزين مؤقت على الخادم لتقليل اختلاف البيانات بين المجموعات، **المستوى الثاني** :تحسين توزيع المعلمات بين العملاء لتعزيز الأداء، وكانت النتائج هي تحسن 15.8% كحد أقصى و 12.6%في المتوسط في الدقة.

الفوائد والتطبيقات

- يحسن أداء التعلم الموحد في بيئات Non-IID، مثل تحليل سلوك المستخدم عبر أجهزة مختلفة.
- يحافظ على الخصوصية مع تحسين الدقة والسرعة.
- مناسب للتطبيقات الواقعية التي تعتمد على بيانات موزعة (مثل الهواتف الذكية، أنظمة التوصيات).

التشابه الوحيد بين هذا البحث ومشروعنا، هو نوع البيانات التي يقوم بمعالجتها وهي غير موحدة المصدر from different distributor ويمتاز هذا النوع من البيانات بالعشوائية وانعدام انحيازه

وقام البحث [3] بتصميم نموذج Hi-LANDER للتجميع الهرمي باستخدام الشبكات العصبية البيانية(GNNs) وهو شبكة عصبية بيانية هرمية (Hierarchical GNN)لتجميع مجموعة من الصور إلى عناقيد (مجموعات) غير معروفة العدد، باستخدام بيانات تدريب مُعلَّمة بعناقيد مختلفة عن تلك المراد تجميعها، كما أنه يعتمد على دمج المكونات المتصلة (connected components)تلقائيًا عبر مستويات هرمية متعددة، ويستفيد من الإشراف (supervision)في بيانات التدريب لتحديد معايير التجميع، بعكس الأساليب غير التقليدية unsupervised hierarchical clustering، نتيجة هذا البحث هو زيادة 49%في F-scoreمقارنة بأساليب التجميع القائمة على GNN، وأبدى تحسن 7%في مقياس NMI (Normalized Mutual Information)، بالنسبة لسرعة الأداء فقام توحيد نموذج التنبؤ باحتمالات الربط (linkage) وكثافة العقد (node densities) في إطار واحد، مما يقلل التكلفة الحسابية بثلاثة أضعاف مقارنة بالطرق الأخرى، ميزاته أنه يعمل مع عدد غير معروف من العناقيد، مما يجعله مناسبًا للتطبيقات الواقعية مثل التعرف على الوجوه أو تحليل سلوك المستخدم.

الاستخدامات المحتملة

- تحليل مجموعات الصور الشخصية
- التعرف على الوجوه أو الأشياء في مجموعات غير مُصنَّفة مسبقًا.
- تطبيقات الرؤية الحاسوبية التي تتطلب تجميعًا دقيقًا وقابلًا للتكيف.

وهو البحث الأقرب على ما تم إنجازه في هذا المشروع من ناحية نوع مجموعة البيانات الهرمية وتقارب نماذج التدريب، بالمقارنة معه نستنتج من .Table 2 تفوق نتائج مشروعنا وذلك بسبب الاخذ بأكثر من ميزة بعين الاعتبار

اسم البحث	البحث	مشروعنا VisionAI
طريقة عنونة البيانات	يدويا	Vision-Language model Image Preprocessing CNN models
ميزات مجموعة البيانات	Date – Objects – Folder Name – Image Vector	Date – Objects – Folder Name – Image Vector – Context – Faces – OCR – Background Class – Color Palette
نوع مجموعة البيانات	Hierarchical data – biased	High dimensional – hierarchical data – from different distribution
نماذج التدريب	Cross-modal retrieval between images and behavioral tags	Classification Machine Learning Random Forest Regressor
نوع التدريب	unsupervised hierarchical clustering	Supervised classifier
F1-score	49	96

Table 2: يوضح هذا الجدول مقارنة بين البحث 2 وما تم إنجازه في مشروعنا

المراجع

[1] https://www.geeksforgeeks.org/machine-learning/image-segmentation-using-k-means-clustering/

[2] Zhang, J.; Li, Z. A Clustered Federated Learning Method of User Behavior Analysis Based on Non-IID Data. *Electronics* **2023**, *12*, 1660. https://doi.org/10.3390/electronics12071660

[3] Yifan Xing and Tong He and Tianjun Xiao and Richard Wang and Yuanjun Xiong and Wei Xia and David Wipf and Zheng Zhang and Stefano Soatto “Learning hierarchical graph neural networks for image clustering” year 2021 https://www.amazon.science/publications/learning-hierarchical-graph-neural-networks-for-image-clustering