SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

معسكر سدايا
Data Science

مشروع

تحليل تغريدات اللغة العربية

من خلال تحليلها وتصنيفها

فريق العمل

رهف العيسى / داليا العنزي / أمل الشهابي

مـــبـــادرة

ســـدايـــا

مـن ضـمن مـــبادرات سـدايـا الـتي تـسعى مـن خـلالـها إلـى الـتركـيز عـلى مـعالـجة الـبيانـات وتـقديـم خـدمـات وحـلول تـساهـم في تـطبيق أعـلى اسـتفادة مـن الـبيانـات و أدوات الـذكـاء الإصـطناعـي، ومـن ضـمن هـذه المـبادرات مـبادرة الـتركـيز عـلى تـقديـم حـلول فـي مـعالـجة الـبيانـات المـتعلقة بـالـلغة الـعربـية، ومـن هـذا المـنطلق سـعينا مـن خـلال دورنـا كـعلماء بـيانـات عـلى تسـليط الـضوء عـلى الـلغة الـعربـية.

صـــورة مـن مـــنشور المـــبـــادرة *

فـكـرة

المـــشـــروع

في طور عـملنا عـلى هـذا المشـروع بـحثنا عـن أكـثر المـصادر الـتي تـحتوي عـلى الـنص الـعربـي فـي مـواقـع الـتواصـل الإجتماعـي و كـان أبـرز تـلك المـنصات هـي مـنصة **تـويـتر** الـتي تـحتوي عـلى مـلايـين المـنشورات بـالـلغة الـعربـية فـي عـدة مـجالات مـتنوعـة، ومـن هـذا المـنطلق قـررنـا الـعمل عـلى الاسـتفادة مـن تـلك الـبيانـات مـن خـلال سـحبها و تحـليلها ومـن ثـم تـصنيفها بـاسـتخدام أدوات الـذكـاء الإصـطناعـي.

مالذي قمنا بعمله؟

البداية

مرحلة سحب البيانات

في هذه المرحلة قمنا بسحب البيانات من خلال تويتر API

مرحلة معالجة البيانات

قمنا في هذه المرحلة بعدة مراحل لمعالجة البيانات وتحليلها

مرحلة استعراض البيانات

في عده المرحلة قمنا باستعراض البيانات عن طريق الرسم البياني

مالذي قمنا بعمله؟

الإطلاق

مرحلة بناء لوحة البيانات

في هذه المرحلة قمنا بعمل
داشبورد توضيحي يعكس نتائج
التي توصلنا لها في تحليل البيانات

مرحلة التحقق من النتائج

قمنا في هذه المرحلة بعملية اختبار
للتحقق من نتائج المودل المختار

Model مرحلة بناء ال

في هذه المرحلة قمنا بتجربة أكثر
من نوع للمودل تعلم الآلة

مرحلة سحب البيانات

١ سحب الداتا قانونيًا

رفع طلب حساب المطورين (developer account)

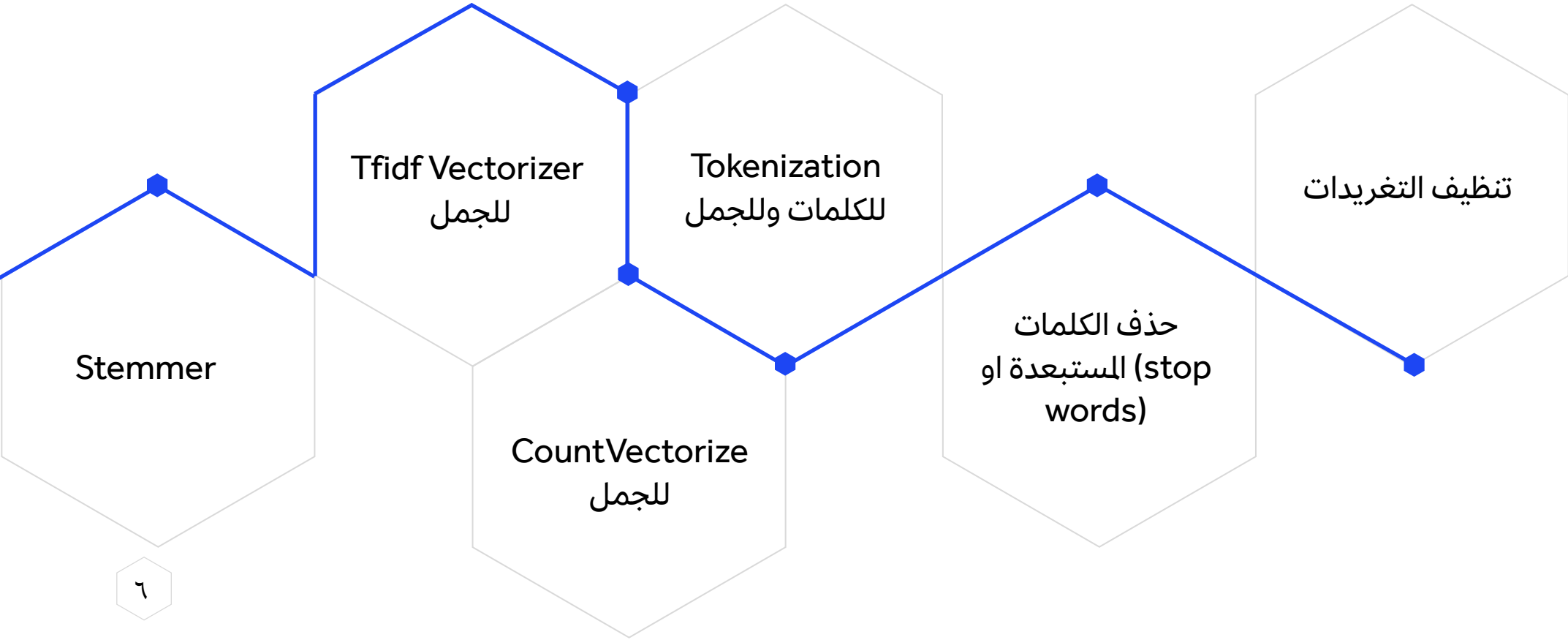
٢ استخدام وتوليد API’s keys

البدأ بالكودينق

٣ تصفية محتوى التغريدات

٤ حفظ الداتا في ملف اكسل

مرحلة معالجة البيانات



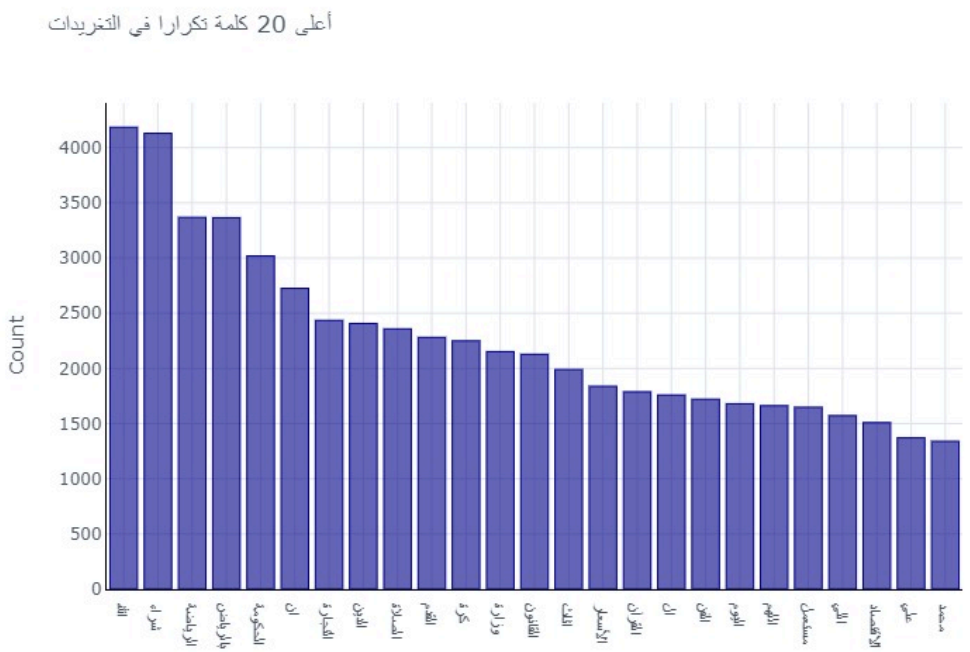
مرحلة استعراض البيانات

EDA

١- رسم بياني يستعرض في تصنيفات التغريدات.

مرحلة استعراض البيانات

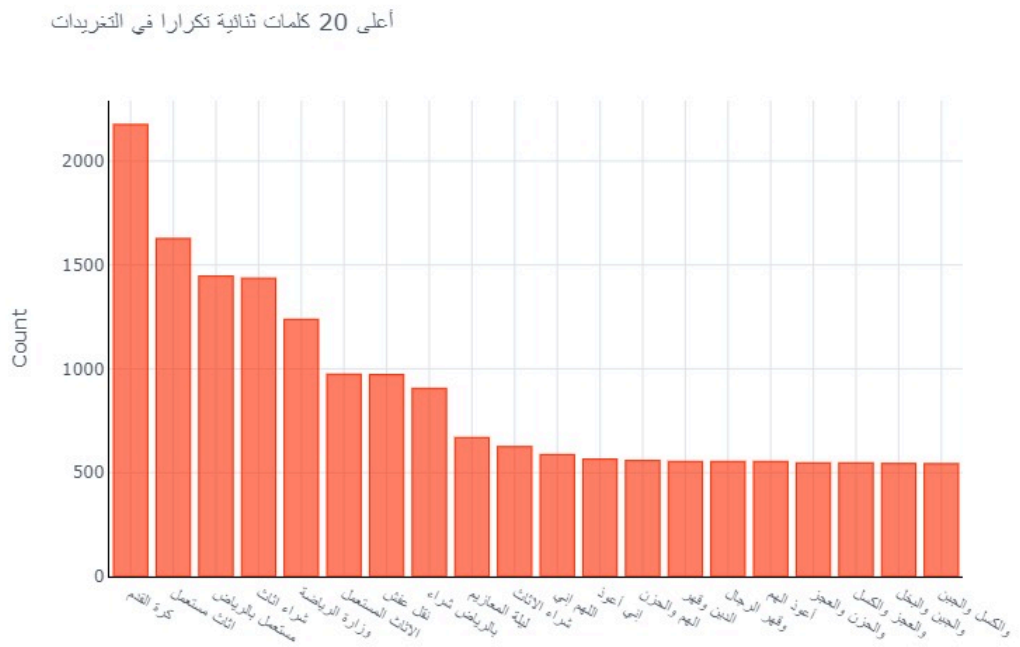
EDA



٢- رسم بياني يستعرض فيه أكثرالكلمات تكرارا في التغريدات.

مرحلة استعراض البيانات

EDA



٣- رسم بياني يستعرض فيه أكثر الكلمات تكرارا في التغريدات.

مرحلة استعراض البيانات

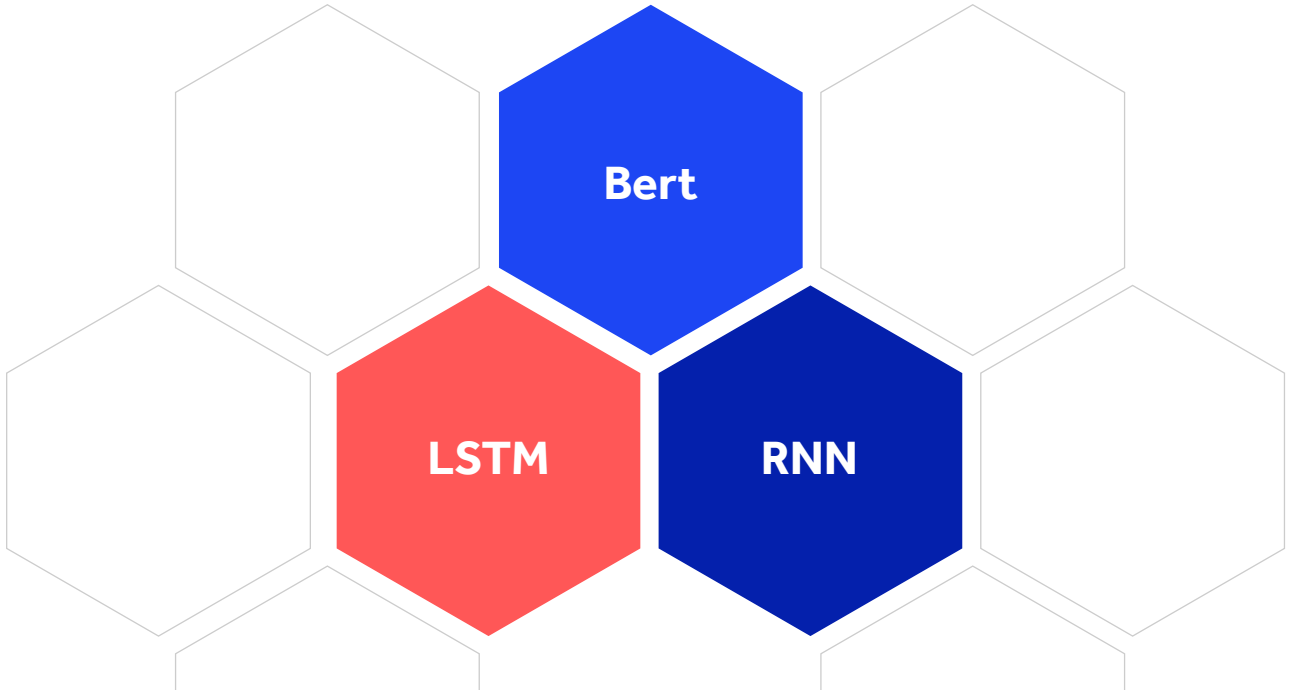
EDA

٤- رسم بياني يستعرض فيه أكثر الكلمات تكرارا في التغريدات.

مرحلة استعراض البيانات EDA

٤- رسم بياني يستعرض فيه تحليل المشاعر .

مرحلة بناء Model



مودل Bert

Report Bert Model لتحليل المشاعر

	precision	recall	f1-score	support
0	0.77	0.82	0.80	4514
1	0.81	0.76	0.79	4535
accuracy			0.79	9049
macro avg	0.79	0.79	0.79	9049
weighted avg	0.79	0.79	0.79	9049

```
array([[3724,  790],
       [1083, 3452]])
```

مودل Bert

Report Bert Model للتصنيفات

	precision	recall	f1-score	support
0	0.94	0.95	0.95	1891
1	0.96	0.95	0.95	1803
2	0.93	0.93	0.93	1894
3	0.95	0.95	0.95	1855
4	0.96	0.97	0.96	1974
accuracy			0.95	9417
macro avg	0.95	0.95	0.95	9417
weighted avg	0.95	0.95	0.95	9417

```
array([[1796,  11,  27,  33,  24],
       [ 23, 1708,  42,  21,   9],
       [ 26,  43, 1754,  37,  34],
       [ 28,  10,  36, 1768,  13],
       [ 29,   7,  17,   8, 1913]])
```

مودل LSTM

LSTM Model للتصنيفات

Accuracy: 91%

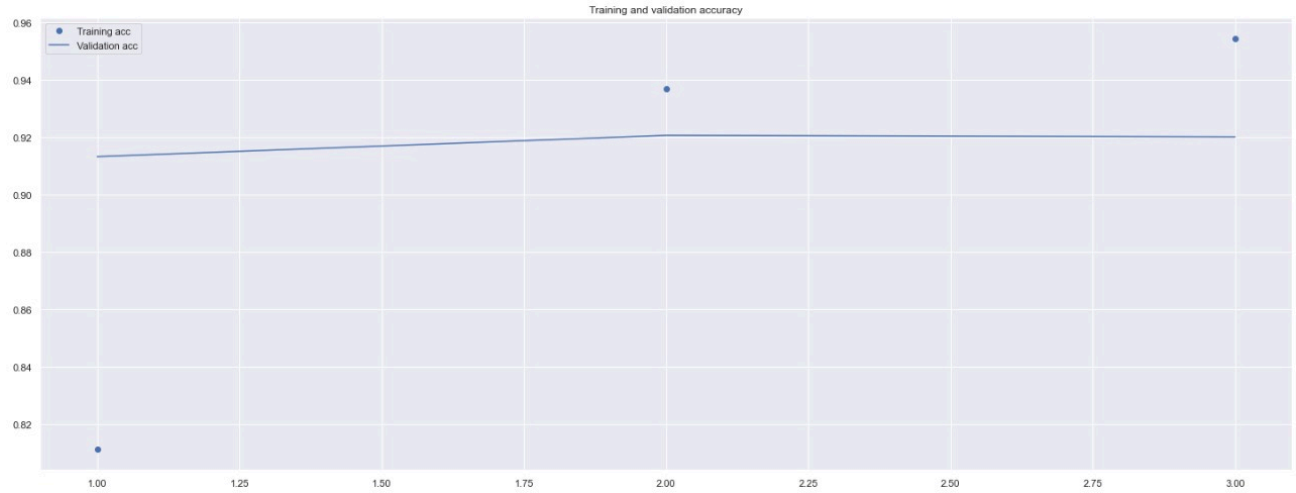
مودل RNN

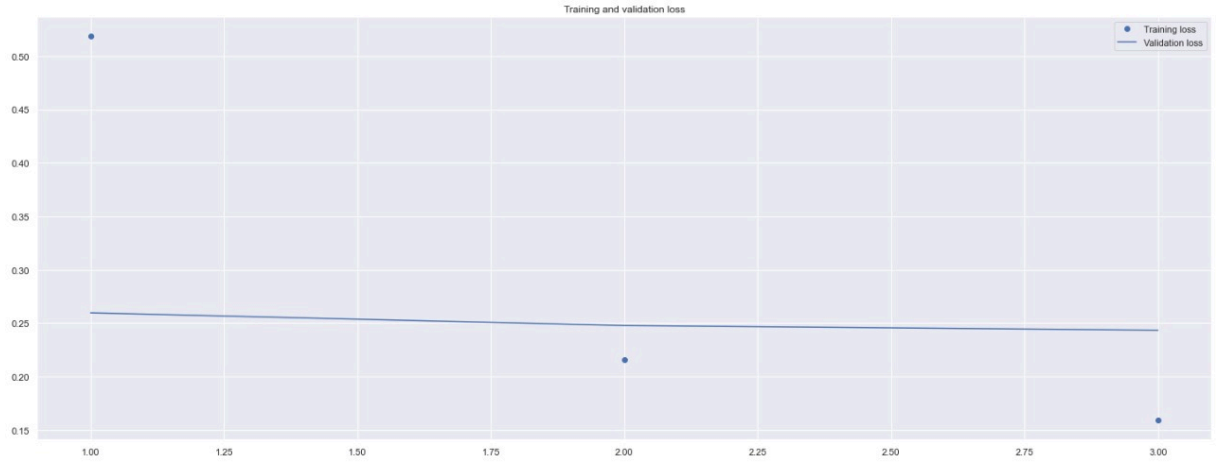
RNN Model للتصنيفات

Accuracy: 20%

مودل LSTM

رسم بياني لمودل LSTM

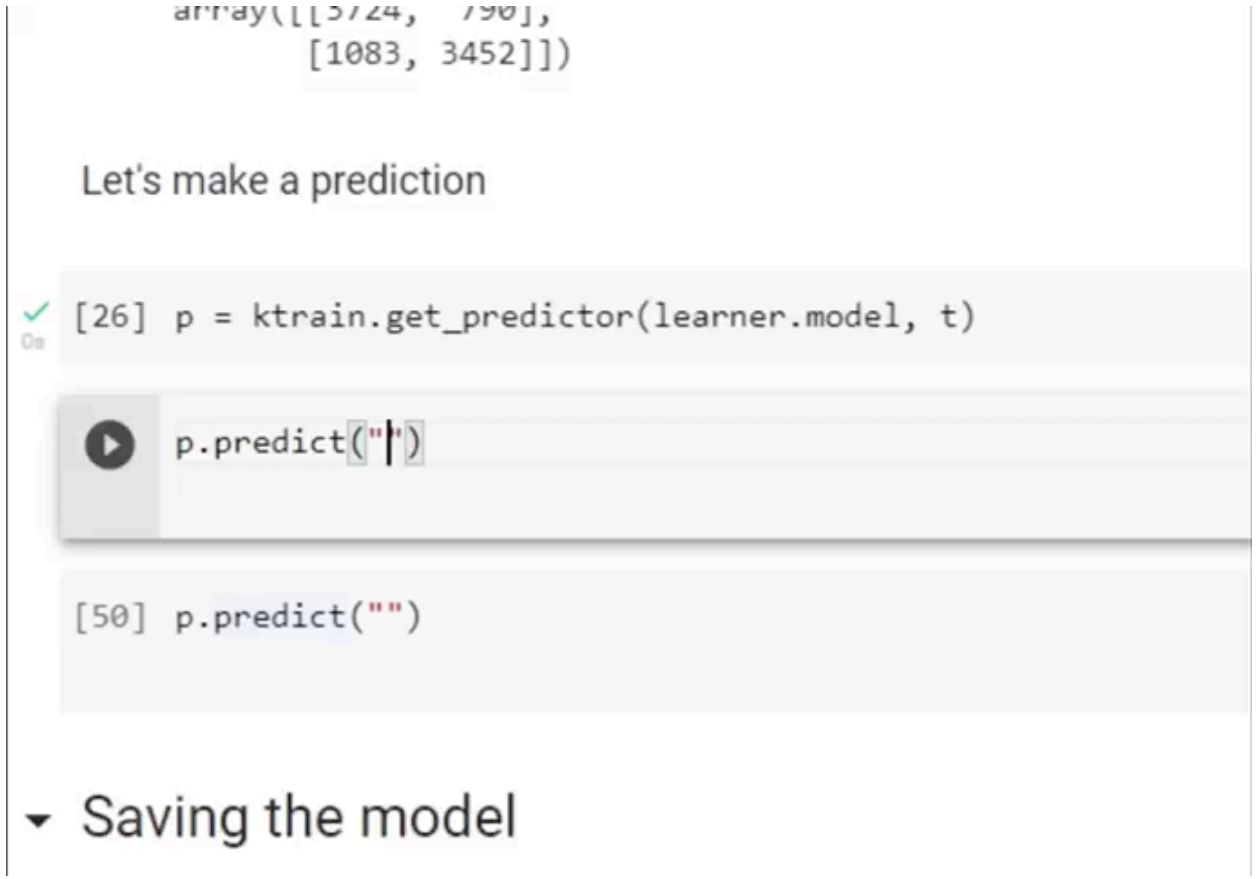




مرحلة التحقق من النتائج

Demo

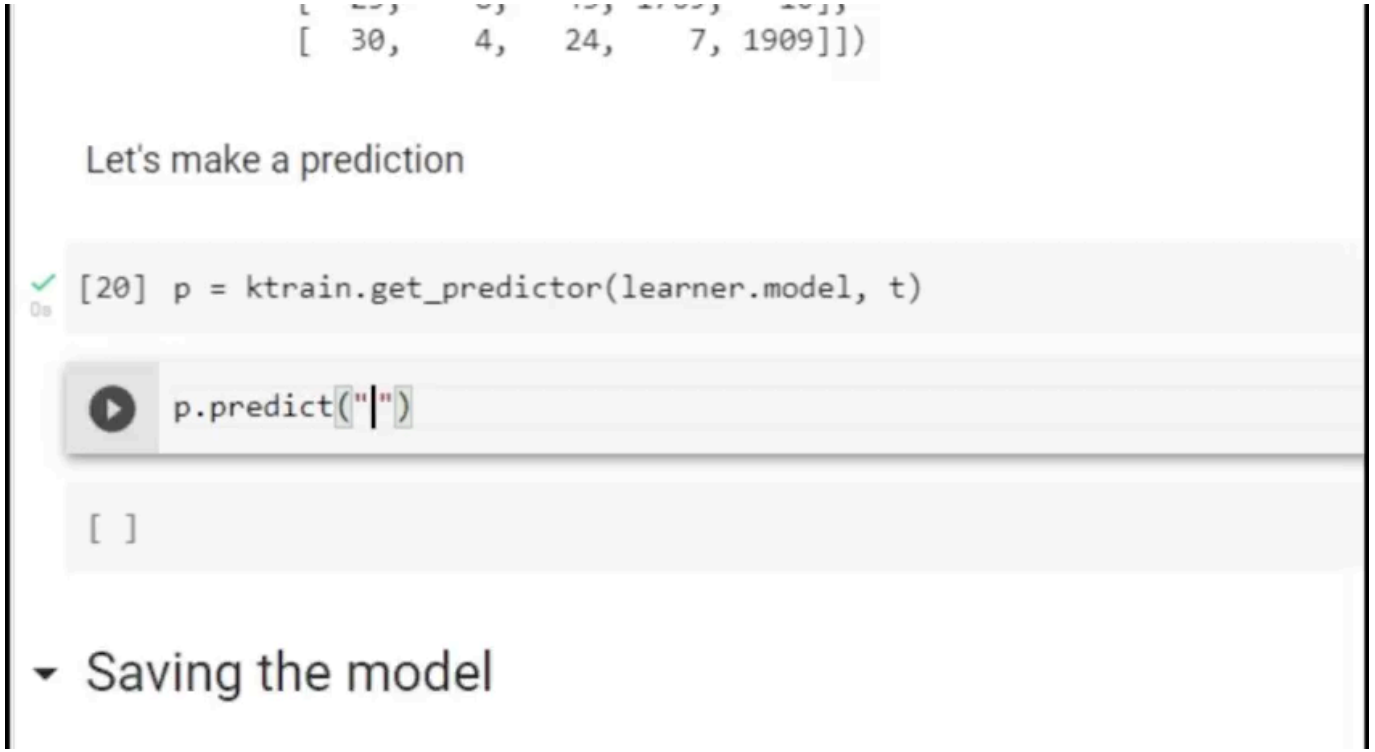
فديو توضيحي على تجربة المودل والتحقق منه
في تحليل المشاعر



مرحلة التحقق من النتائج

Demo

فديو توضيحي على تجربة المودل والتحقق منه في التصنيف



التحديات التي واجهتنا

في المشروع وقدرنا على التغلب عليها

- تحدي سحب الداتا
- التعامل مع اللغة العربية وقلة المراجع
- تعامل ديب ليرنينق مودل مع اللغة والعربية والمشاكل

مرحلة بناء لوحة البيانات

الختامية

الحمدلله على تمام المشروع بنجاح

في العمل المستقبل نتمنى تطوير وأنشاء موقع خاص في تحليل وتصنيف اراء ومشاعر الناس بالنص العربي

شـــكرا لـــكم

نـتمنا ان يـكون نـال اعـجابـكم مشـروعـنا
ومـا تـم طـرحـه

فريق الـعمل

رهف العيسى / داليا العنزي / أمل الشهابي