



# NATURAL LANGUAGE PROCESSES

Arabic tweets analysis



**30 Dec 2021**



Rahaf Alessa



Dalya Alanazi



Inshirah Almutairi

---

Dr.Mejdal Alqahtani

# Work plan



Step 01



-Introduction  
for dataset

Step 02



-Fetch the data  
using scraping

Step 03



-Data Pre-processing

Here you could  
describe  
the topic of the  
section

Step 04



- EDA

display chart for more  
visulazation

Step 05



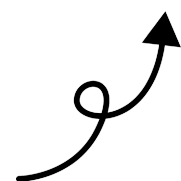
- Model Selection

Here you could  
describe the topic of  
the section the topic  
of the section

Step 02



- Conclusion



WE ARE HERE!



# 01- Introduction

 Our aim is to analysis Arabic recent tweets.



# من الألف إلى الياء

نطور، ونستحدث، ونبتكر

## • طموحنا •

أن تكون المملكة العربية السعودية الرائدة عالمياً في  
مجال معالجة اللغة العربية بالذكاء الاصطناعي

### وضعت سدأيا خدمة اللغة العربية كأحدى أولوياتها وذلك عبر:

تأمين حركتها التميز لمعالجة اللغة العربية  
والذكاء الاصطناعي بالشراكة مع عدد من الجهات  
الوطنية ذات العلاقة والمؤسسات الأكاديمية  
والشركات العالمية الرائدة في هذا المجال.

تطوير إستراتيجية معالجة اللغة العربية بتقنيات  
الذكاء الاصطناعي، التي تضمنت أبحاثاً ومبادرات  
إستراتيجية لتتواءم المملكة الرائدة في المجال.

تطوير نماذج للتصنيف النحوي على المستوى  
والفهم العربي واستخدامها في تطبيقات الرؤية  
الاصطناعية.

تطوير نموذج متقدم للتعريف التلقائي على  
النظام باللغة العربية مع التركيز على الفهم  
السويدي، وذلك بتسوية لغة النسخ إلى  
العام.

### عن معالجة اللغة الطبيعية:

أحد معالجة اللغة الطبيعية هو ذلك الذي يعنى بعمل الآلة في إدارة  
على فهم وتوليد اللغة مثل البشر.  
ويهدف أن تقوم الآلة بما نقوله ونفكره، ومعالجتها بعد ذلك بناء تطبيقات اللغة من طريق استخلاص المعنى  
والبيانات من المدخلات الصوتية أو النصية.



## 02- Fetch the data

 We requested a developer account and got it to fetch 29,000 tweets in Arabic



# 03-Data Pre-processing

*pre-p steps*



 - Tokenaize

 - Stemming

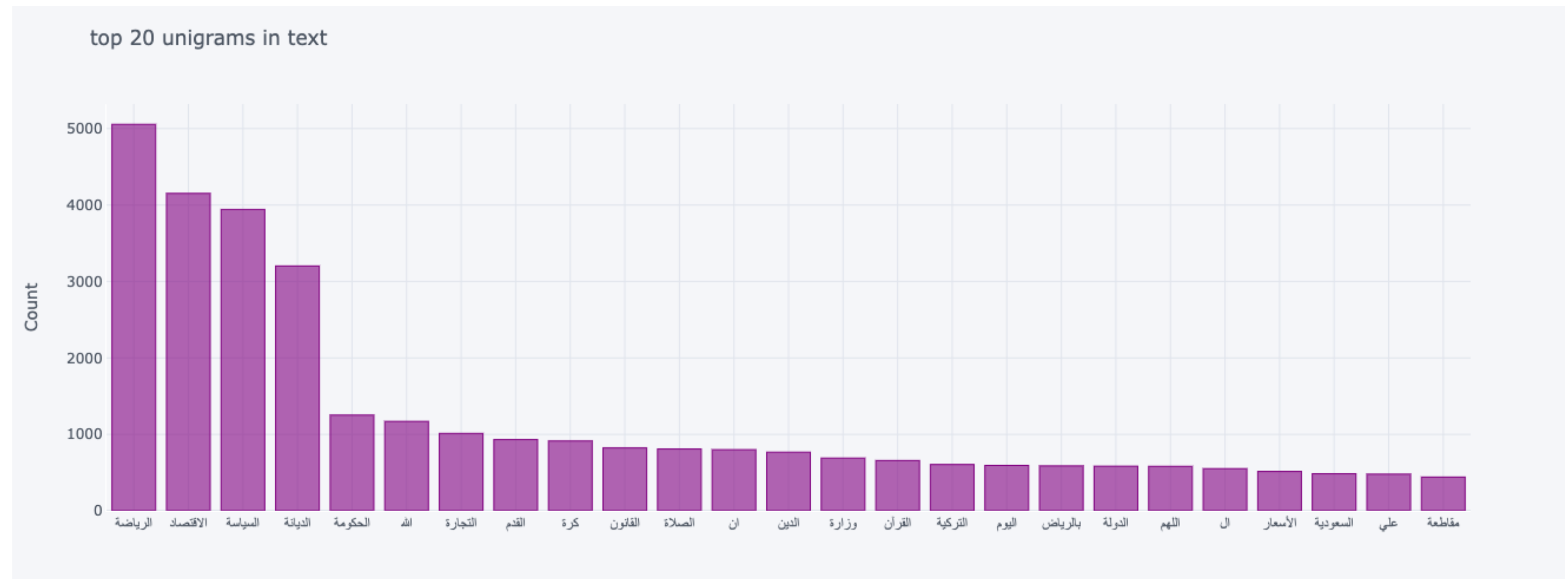
 - Stopwords

 - CountVecatorizer





- This chart is represent the most 20 words.





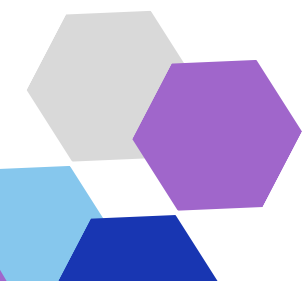


## 06-Topic modeling

---

 1- Latent Semantic Analysis(LSA)

 2- Non-negative Matrix Factorization(NMF)





# 06-Topic modeling

## 1- Latent Semantic Analysis(LSA)

### Topic 0

التركية الدولة , رسمي الإجابة , معاهدة التجارة , موقعة معاهدة , التجارة الدولية , قرار رسمي , يسألون , معاهدة , الإجابة , موقعة , تصدر قرار

### Topic 1

ة سيساهم , العثمانية البائدة , اسقاط الاقتصاد , رحمهم اسقاط , اسقاط الدولة , الادو الاقتصاد , الادو , اجدادنا رحمهم , التركي المترنج , اختلفت

### Topic 2

ج , مايسمج , متخصص رياضة , الجيم نصيحة , هالحساب ويستفيد , هالحساب , ويستفيد متخصص , ويستفيد , ويقدر , نصيحة يمارس , منزلية , الرياضة ويقدر

### Topic 3

ني يصلي , اجعلني , القرآن الكريم , الكريم , ثلث , ثلث القرآن , الكريم رتويت , رتويت لعلها , لعلها , تشفع , لعلها تشفع , تشفع الديانة , رتويت



# 06-Topic modeling

## 2- Non-negative Matrix Factorization(NMF)

Topic 0

حساب, ان, وزير, شكرا, الاتحاد, نادي, الي, رياضة, رئيس, الاهلي, علي, البيت, يكون, وفقا, حسين, الدوري, يمارس, يتابع, نصيحة

Topic 1

, قرار, شركات, البضائع, الدولية, تصدر, الأسعار, الإجابة, اكبر, رسمي, للسعودية, موقعة, معاهدة, خلاص, يسألون, ارسلت, تطالبها

Topic 2

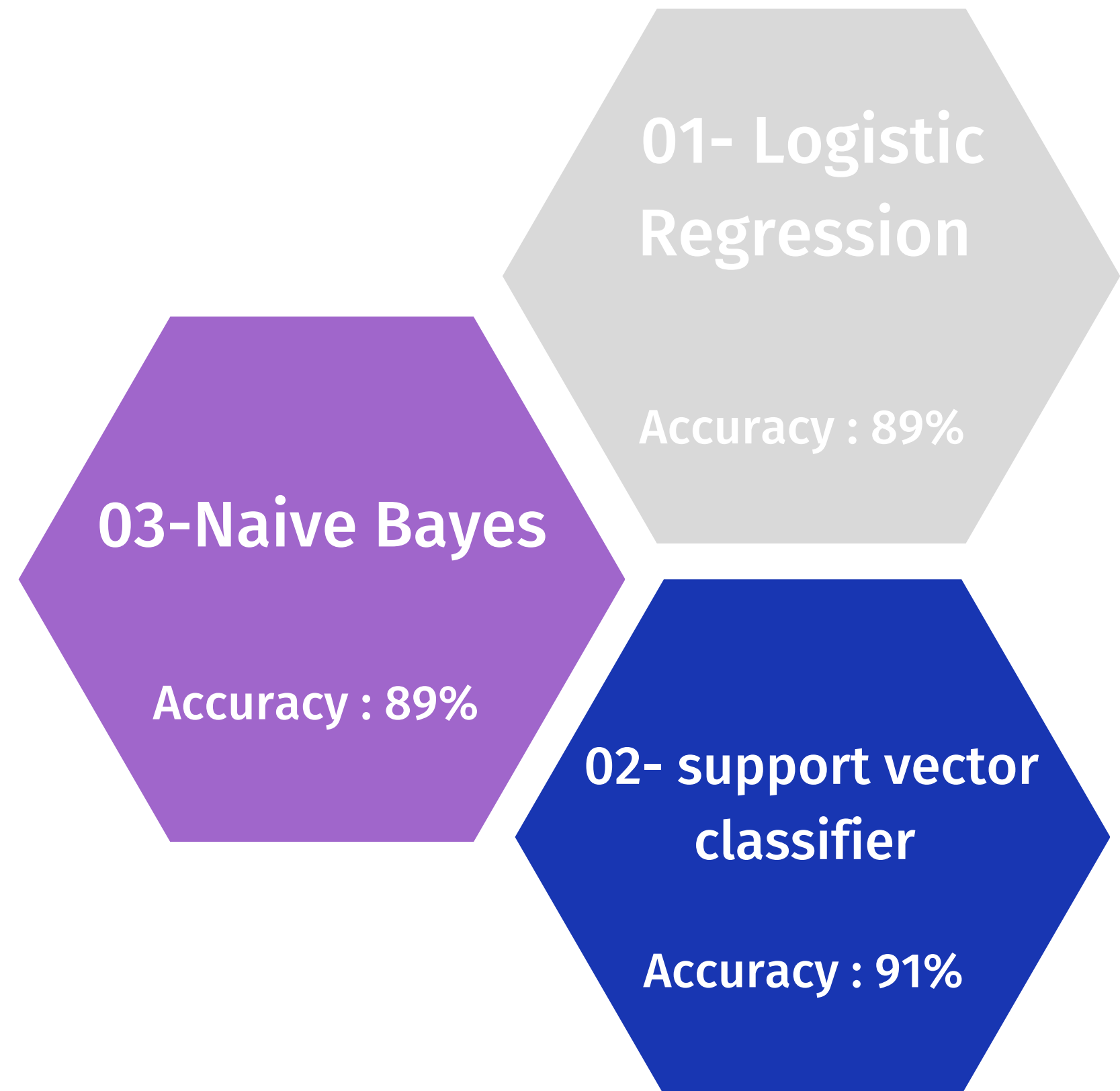
, مواطن, بمناسبة, رئيس, تصرف, المساعدة, السعودية, كنوع, أنواع, بالمملكة, قانون, الشعب, الى, الحريري, القطاع, حدد, موقفك

Topic 3

د, الهم, أعوذ, والكسل, والبخل, وقهر, والعجز, والجبن, الجمعة, صل, النبي, وغلبة, الكريم, علي, يوم, والسلام, ليرتاح, ان, بن



## 05- **Model selection**





- [illegible]





-



- [illegible]



- اليهود الديانة تلت الصلاة والعبادة  
 الصلاة والسلام والعجز والكسل الدين وقهر  
 إني أعوذ الديانة  
 الهم والحزن والعجز الديانة  
 القرآن الكريم وقهر الرجال

## 07-Tools



jupyter



Colab



Pandas



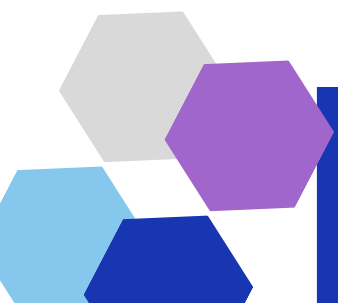
*NumPy*

NumPy



NLTK

NLTK





## 08- Conclusion

SVC is the best model with the highest accuracy.  
we are looking for adding more clusters and increase accuracy.



SCAN ME



# Women to impact

Arabic sentiment  
analysis challenge





**Thank you for listening**

