



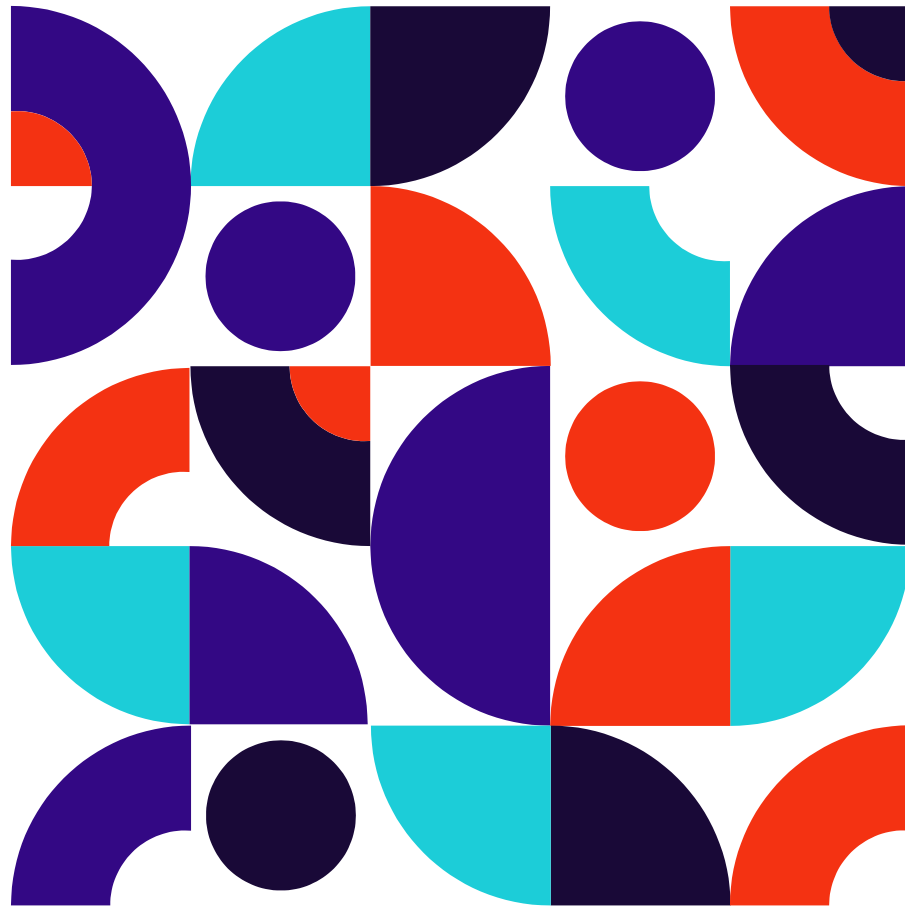
KNA Company

analysis

analysis

KNA Company

Products Review using NLP



Outlines

Introduction

1

Data

2

Topic Modeling

3

Cluster

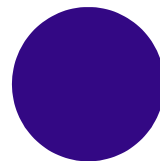
4

Models

5

Summary

6



1

Introduction





Introduction

When bringing any product to the market, it is very important to know the customer review about the product, whether the review is negative or positive, from which the product is evaluated, analyzed, trying to develop the product, etc.

Customer reviews are very important in evaluating any product. We provide a service to analyze reviews, if they are negative or positive, for companies' products, to know the opinions of customers about any product.

2

Data



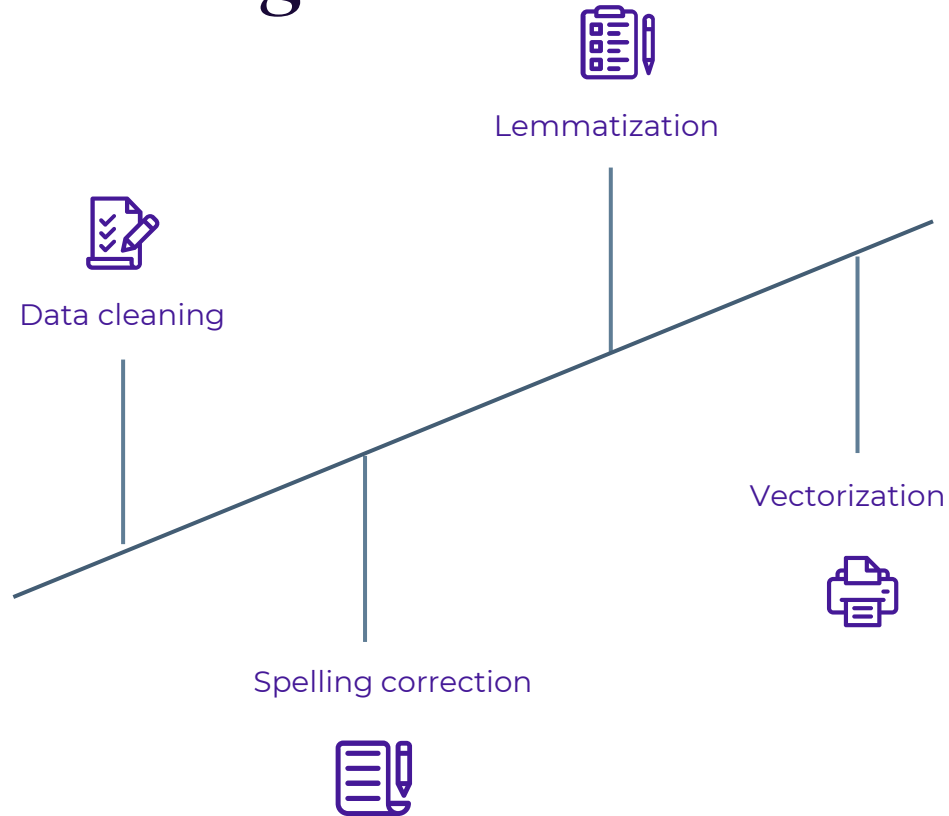


Data

The Data was collecting from Kaggle, it contain the text reviews and label, and it have 100K rows and 2 columns, with TSV file.

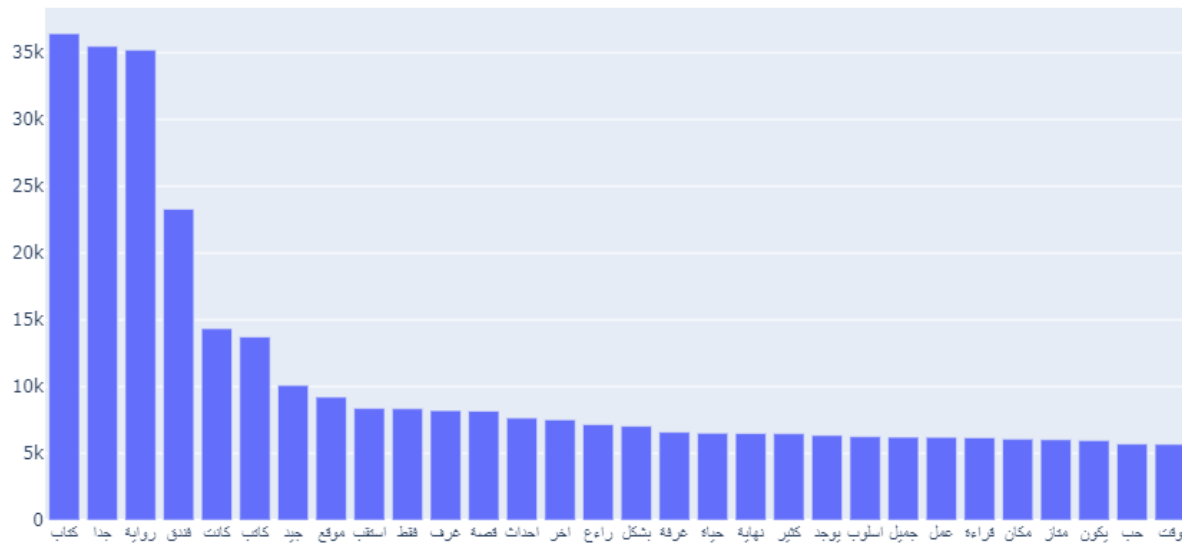
Label represent the sentiment of the customer, its between **Positive, Negative, Mixed.**

Data preprocessing



Data Visualize

Top 30 unigrams in the question text after removing stop words and lemmatization





3

Topic Modeling

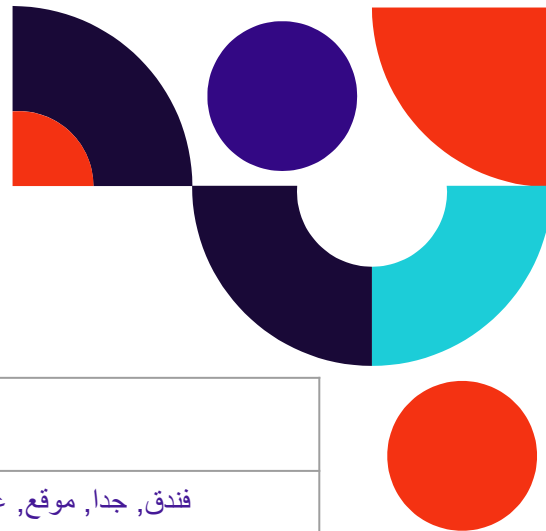


Topic Modeling:



LSA	
Topic 0	رواية, كتاب, جدا, كاتب, كانت, احداث, قصة, اخر, نهاية, حياة, بشكل, قراءة, يكون, كثير, شخصية
Topic 1	رواية, احداث, شخصيات, نهاية, روايات, شخصية, قصة, بطل, لرواية, مراد, كاتبة, وصف, سرد, حبكة, تفاصيل
Topic 2	جدا, فندق, غرف, استقب, غرفة, موقع, جيد, يوجد, سيء, متاز, خدمة, مكان, نظافة, غرفه, فقط
Topic 3	فندق, كانت, انسان, حياة, اذا, حب, اخرى, يكون, عندما, وان, اخر, فقط, يوم, رجل, دين

Topic Modeling:



LDA	
Topic 0	فندق, جدا, موقع, غرف, استقب, جيد, غرفة, يوجد, سيء, ضعيف, نظافة, مكان, فقط, خدمة
Topic 1	كتاب, رواية, كاتب, كانت, حياة, كثير, قراءة, حب, اخرى, اخر, عندما, يكون, وان, انسان
Topic 2	رواية, كتاب, جدا, كاتب, كانت, احداث, روايه, قصة, اسلوب, ده, نهاية, كثير, فكرة, مكن

Topic Modeling:

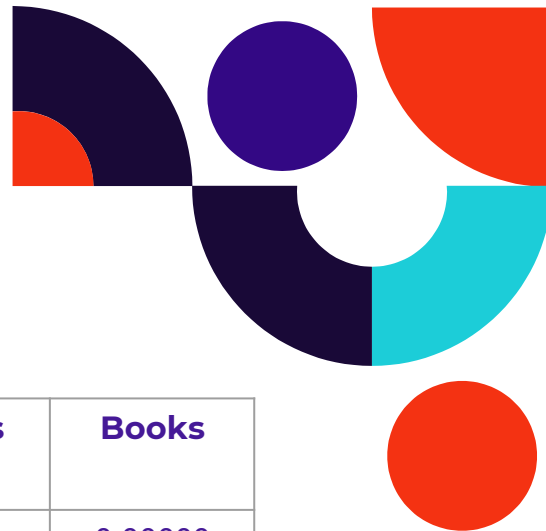


NMF	
Topic 0	رواية, كاتب, كانت, احداث, قصة, نهاية, شخصيات, شخصية, جدا, حياة, اخر, بشكل, حب, فكرة, روايات, اسلوب, عمل, رغم, تكون, قراءة
Topic 1	كتاب, كاتب, كتب, كانت, قراءة, اخر, يكون, كثير, انسان, حياة, دين, جدا, وان, ناس, اسلام, اخرى, فصل, فكرة, بشكل, فقط
Topic 2	جدا, فندق, غرف, استقب, غرفة, موقع, جيد, يوجد, كانت, سيء, فقط, متاز, مكان, خدمة, نظافة, راع, غرفه, ضعيف, وجود, مواقف

Topic Modeling:

NMF

Text	Others	Movies	Books
ممتاز نوعا نظافة وموقع وتجهيز وشاطئ مطعم	0.022105	0.0000	0.00000
ارجع به مرة اخرى قربه بحر مكان قديم توجد خدمات نجوم	0.018286	0.007591	0.003249
كتاب ضعيف جدا ولم استمتع قصه سرد لحه مشهد بدون فكره لقصه	0.00000	0.020182	0.040226



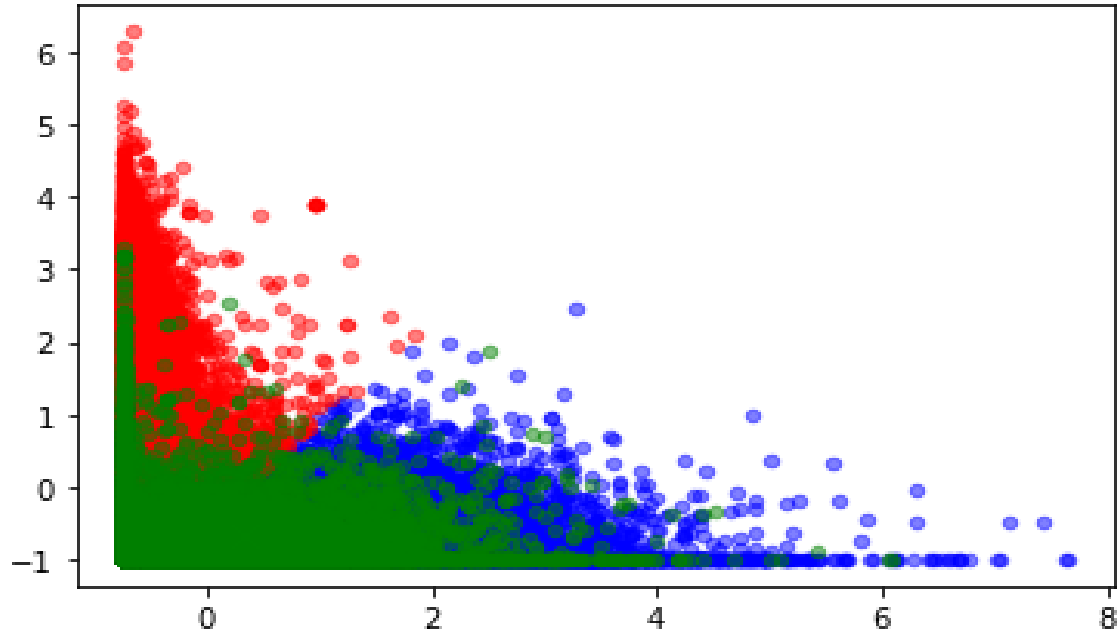
4

Cluster



Cluster

This plot represents the distribution of data after **KMeans** clustering.

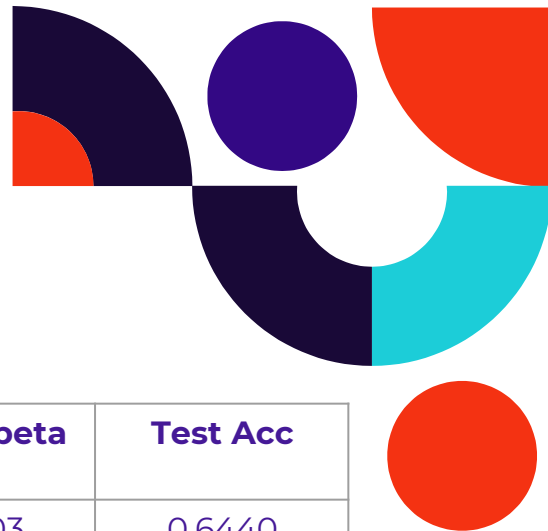


5

Models



Models:



Model	Train Acc	Train Prec	Train Rec	Train Fbeta	Test Acc
Log Reg CV	0.9204	0.9205	0.9204	0.9203	0.6440
Log Reg TFIDF	0.8330	0.8327	0.8329	0.8326	0.6651
NB CV	0.8088	0.8104	0.8088	0.8084	0.6208
NB TFIDF	0.8165	0.8179	0.8165	0.8166	0.6435

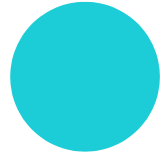
6

Summary





Test case	Results
فلم جميل و اختيار موفق كريستوفر نولان بتحفة فنية جميلة اخرى و تمثنييل ولا اروع من الاسطورة ليوناردو دي كابريو	Positive
كتاب سيئ و رواية مفرعة يوجد بها خروج عن النص بشكل فظيع قد يكون الكتاب الاخير الذي سأشتره لهذا الكاتب	Negative
فندق مقبول المواقف قديمة الاثاث مستحدث انصح به اصحاب الميزانية المحدودة قد يكون خيار جيد	Mixed





Summary

We do the lemmatization and stemming with multi libraries, we decide **qalsadi** libraires is best one.

We found the best NLP practice for our data is **Topic Modeling**, with **NMF**, the **cluster** didn't give us a good result, and for classification model **Logistic Regression** with **TFIDF**.





Thanks !