Tawakkalna App Reviews

get sentiment from reviews

توكلنا

Tawakkalna

By: Fahad Alharbi

TABLE OF CONTENTS

BACKGROUND

01

ABOUT COMPANY

Tawakkalna is application to serve citizens and residents by knowing official documents and performing some government services

توكلنا
Tawakkalna

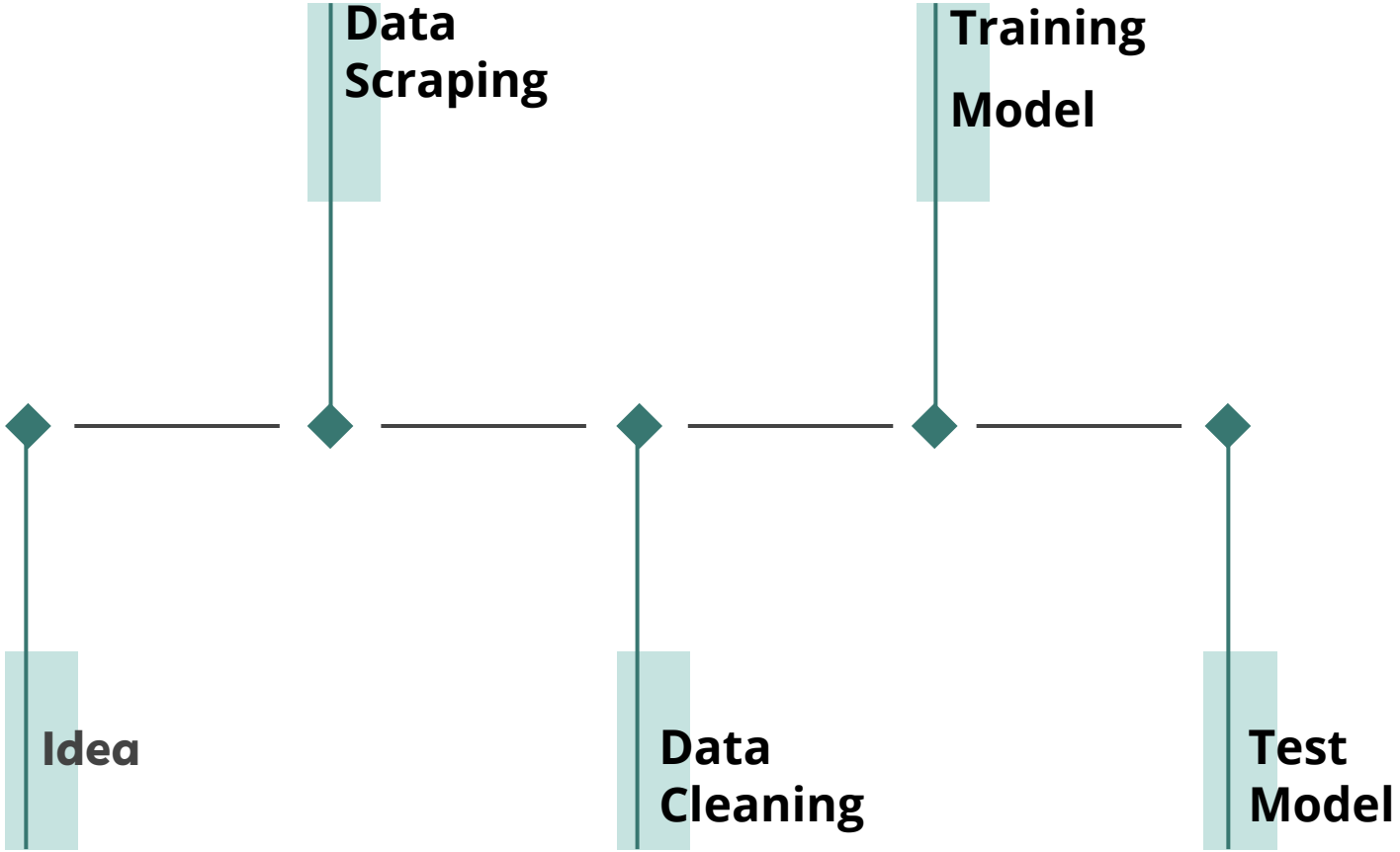
PROBLEM STATEMENT

The target of this project is to know the sentiment from reviews on the Tawakkalna application

METHODOLOGY

02

METHODOLOGY



Data Scraping

App Store

Apple Store

Scraping data reviews for the period from April 2020 to September 2021

Google Play

Google Play Store

Scraping data reviews for the period from April 2020 to September 2021

Data Cleaning

Drop Nulls

Spaces , emojis , numbers

01

Translate to English

There were many languages in the reviews other than English

02

03

Drop columns

There were many unhelpful columns

04

Merge

Merge Apple store dataset and Google store dataset

Data Description

Apple Store

- **Columns = 7**
- **Rows =** title **3112** + reviews **3112** **= 6224**

Google Play Store

- **Columns = 10**
- **Rows = 37179**

After merging two dataset and dropping columns are not needed

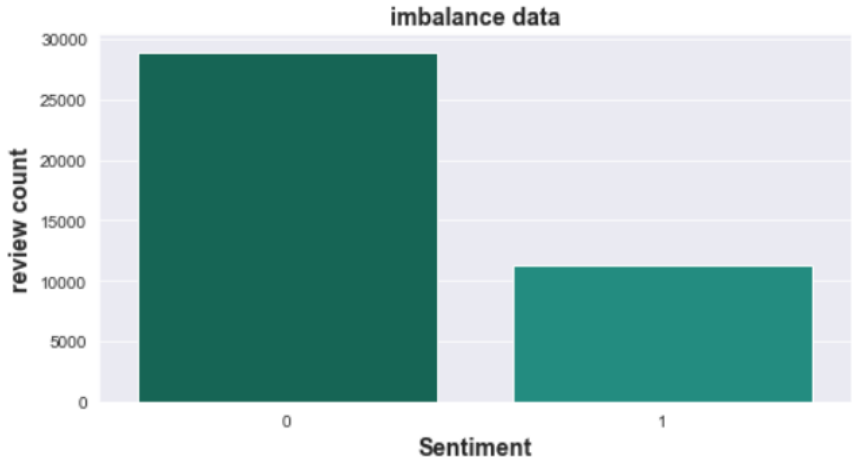
- **Columns = 5**
- **Rows = 43403**

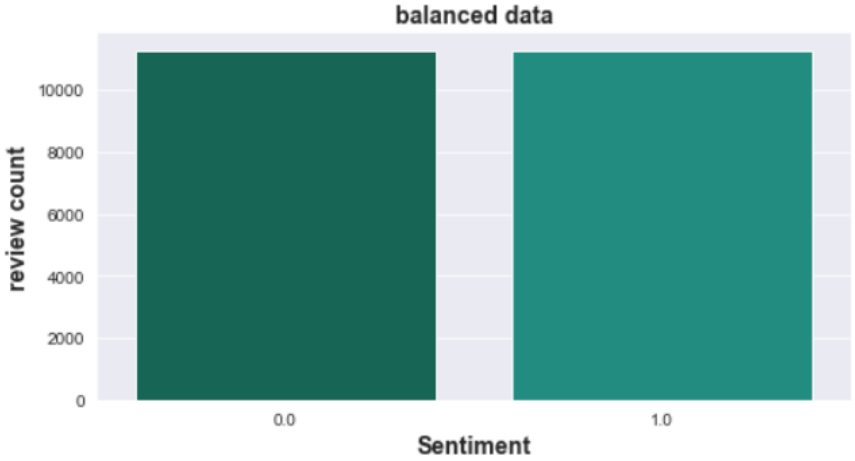
Get Sentiment From Rating

- Above 4 is a good sentiment which I chose to be equal to 0
- Below 2 is a bad sentiment which I chose to be equal to 1

Imbalance Data

Good reviews are more than bad reviews





Extracting Features from Text

TF – IDF

5661 words

Weight?	Feature
+3.551	worst
+3.189	bad
+3.006	privacy
+2.918	update
+2.847	fix
+2.825	updating

Weight?	Feature
-3.214	wonderful
-3.659	thanks
-3.763	useful
-4.031	thank
-4.055	excellent
-4.800	great

Count Vectorizer

tokenize a collection of text documents

Training Model

Logistic Regression

F2 Train = 0.88725
F2 validation = 0.83521
F2 Test = 0.84338

xgboost

F2 Train = 0.86755
F2 validation = 0.85652
F2 Test = 0.86155

Random Forest Classifier

F2 Train = 0.91442
F2 validation = 0.80355
F2 Test = 0.80804

K Neighbors Classifier

Training: 0.84383
validation : 0.80350
Test : 0.80617

Decision Tree Classifier

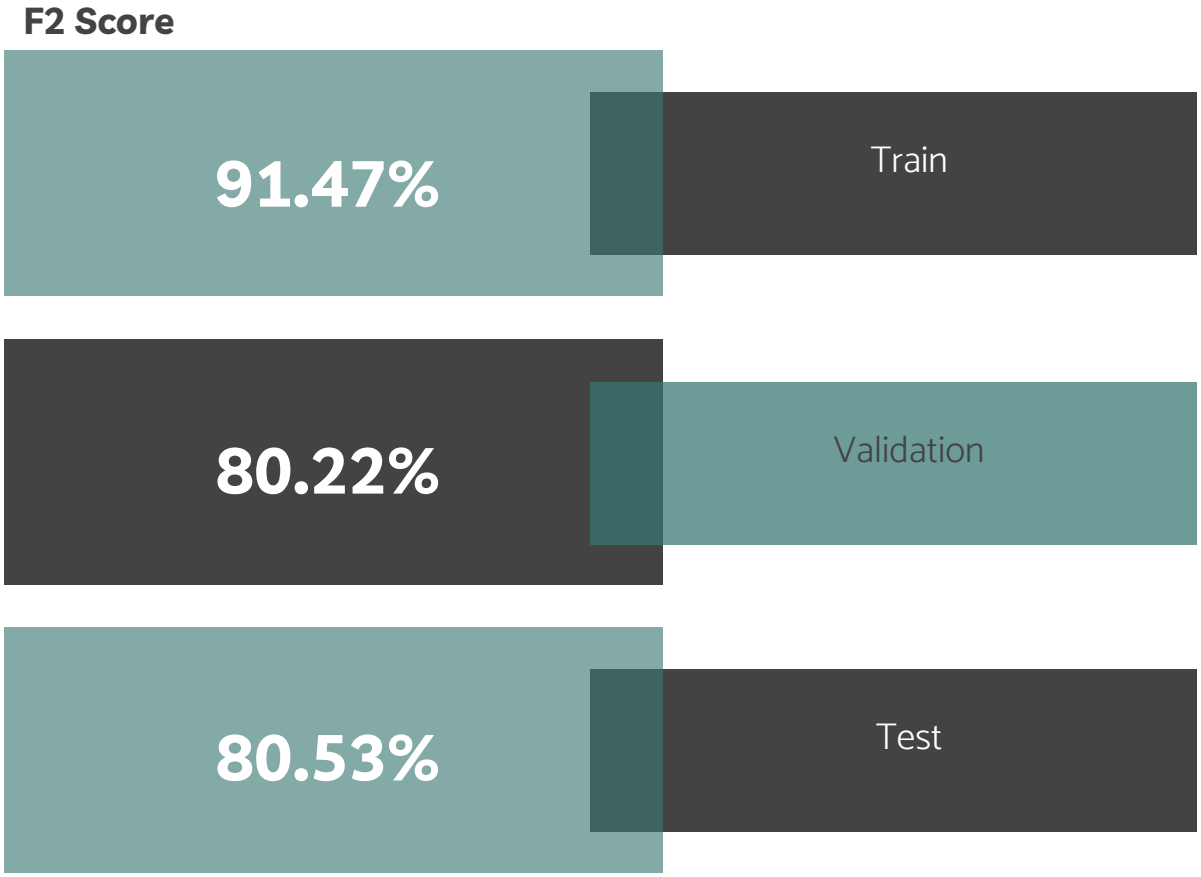
F2 Train = 0.89653
F2 validation = 0.81989
F2 Test = 0.82750

Training Model

Best model: Random Forest Classifier

	Accuracy	precision	recall	F2
Train	93.92%	97.55%	90.065	91.47%
Validation	84.09%	87.95%	78.49%	80.22%
Test	83.43%	87.37%	78.99%	80.53%

Training Model



Test Model

Real Test of The Model

If the review is negative:

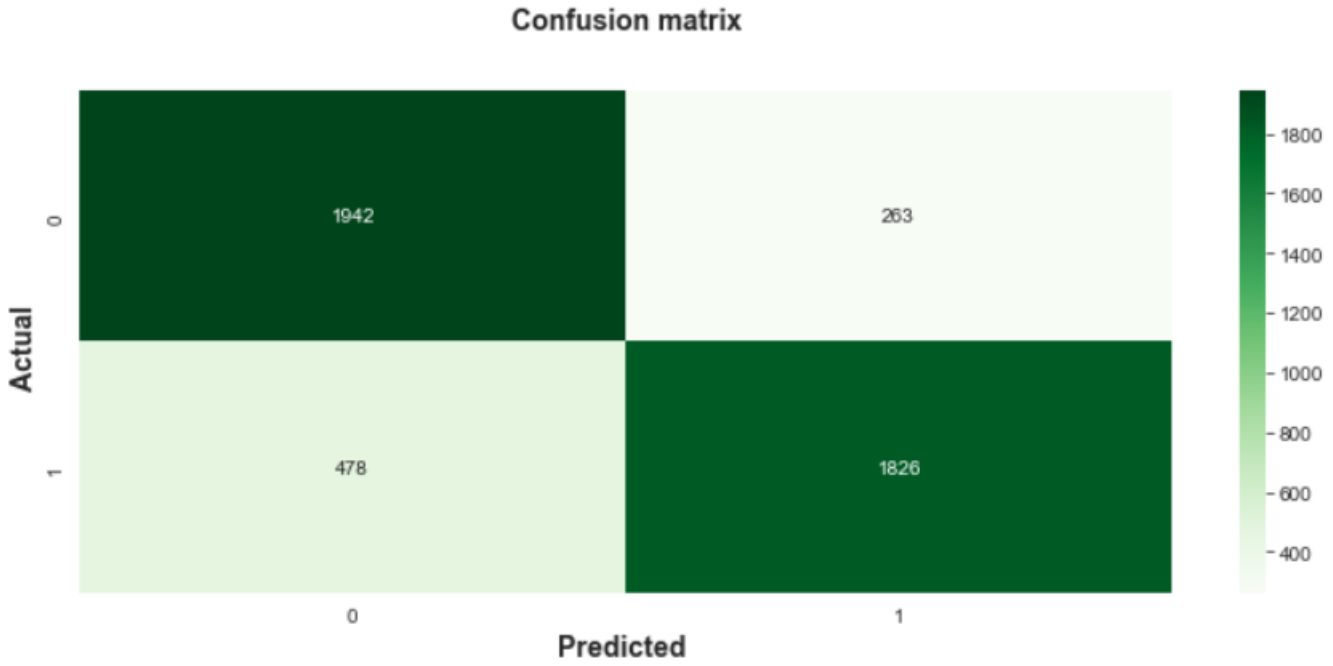
Input : this application needs update

Output : We are very sorry to know that you had a bad experience, we will work to fix the problem

If the review is positive:

Input : this good application

Output : Thank you for the review



More attention to negative comments, which is number 1 in sentiment analysis

Recall Score, F2 Score

CHALLENGES

03

CHALLENGES

Translate

Lots of reviews written in languages other than English and emojis

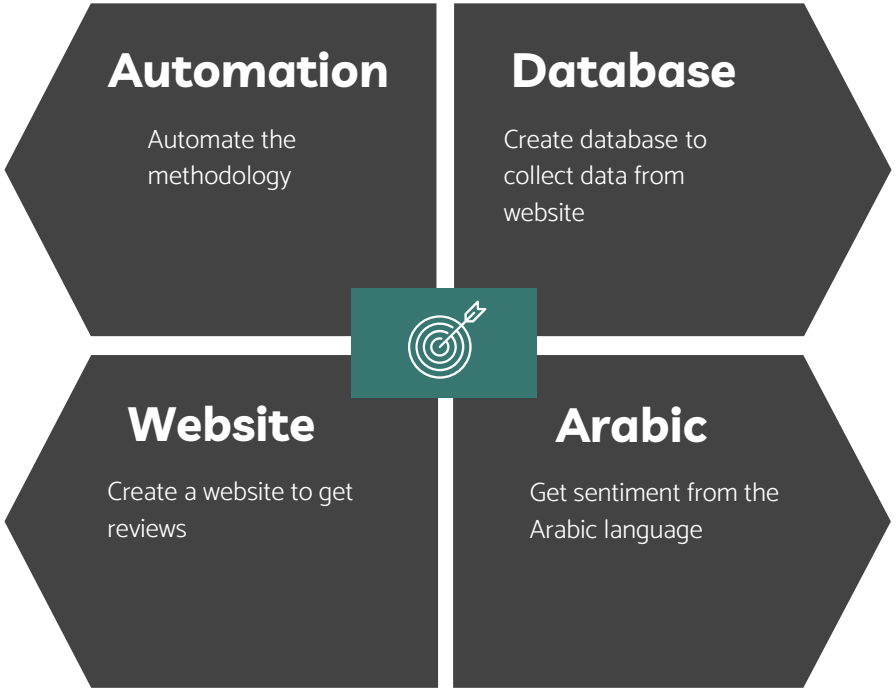
Use new method

TF – IDF method

CONCLUSION

04

FUTURE WORKS



THANKS

Does anyone have any questions?