Cigniti Use-case

Defect Classification

P.Rojarani

Objective

- Build a model which will automate defect manager's job of validating the defects.

- A model should take the new defects as an input and should classify each defect either Accept / Reject (Valid / In-Valid) along with the confidence score for each.

- For an invalid defect, Model should let the user know what could be the possible reason to reject that particular defect.

DataSet Information

1. Valid_defects.json
2. cancelled_defects.json

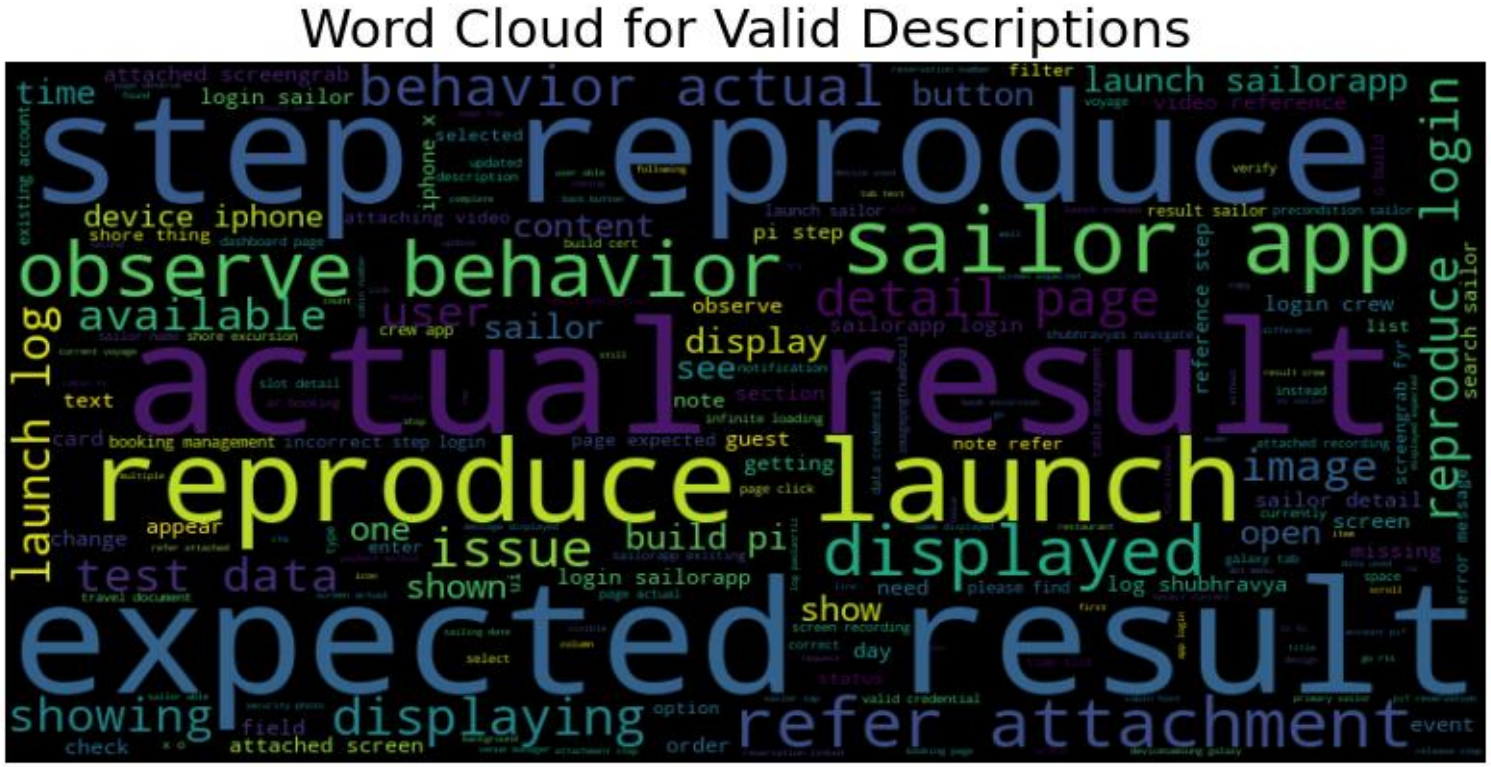
The data dictionary is given below.

1. Status: The status of the defect
2. Description: The user description related to the defect.
3. Summary: Summary related to the defect
4. Comments: The comments related to the defect.
5. Root cause: What is the root cause of the defect
6. Priority: The priority of the defect(High, Medium, Low)
7. Assignee: The developer assigned to the defect.
8. Application: The application name in which the defect was raised.

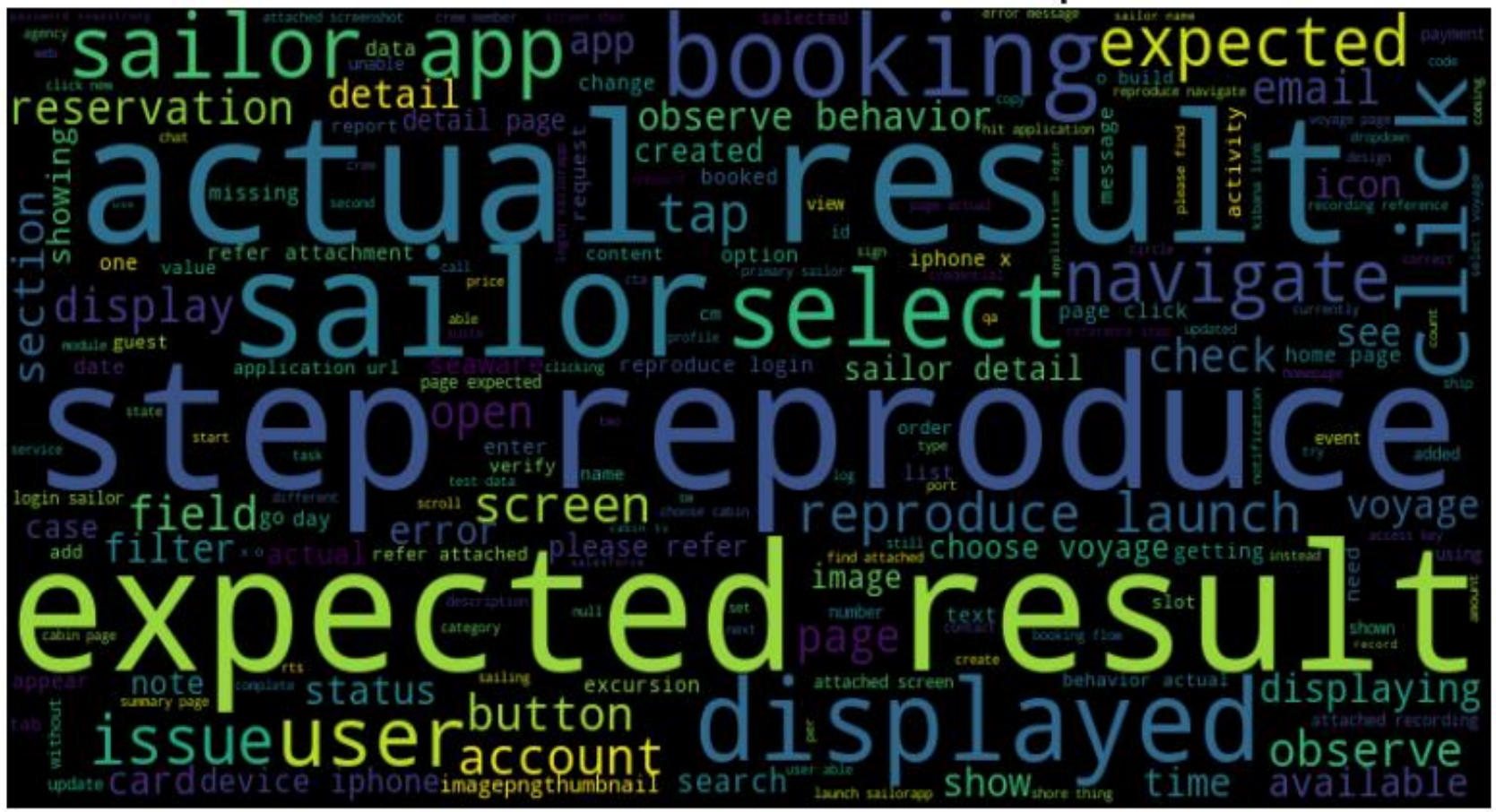
Steps Followed

- Text Normalization
  - *Cleaning texts(removing punc,number etc)*
  - *Tokenization*
  - *Removing stopwords*
  - *Word stemming*
- Converting text into vectors using TF-IDF
- Model building and Evaluation
- Finding the possible reasons for invalid defects using cosine similarity
- Deployement using Streamlit

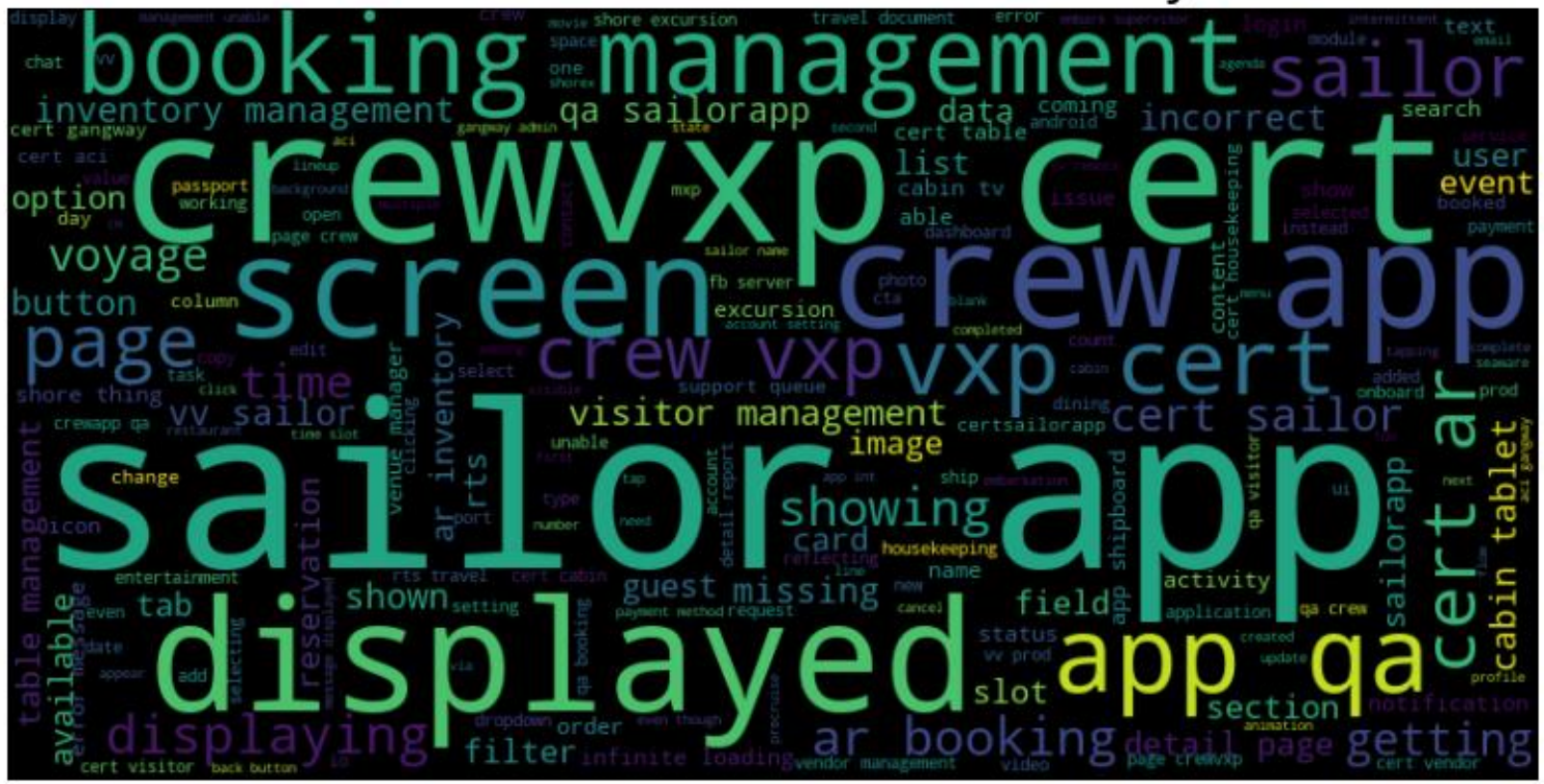
WordClouds


Word Cloud for Valid Descriptions
step reproduce
actual result
expected result
reproduce launch
observe behavior
sailor app
behavior actual
launch sailorapp
reproduce login
launch log
displayed
refer attachment
displaying
showing
test data
issue
build pi
image
open
missing
available
detail page
display
content
device iphone
shore thing
login sailor
time
button
filter
video reference
attached screengrab
attaching video
pi step
result sailor
precondition sailor
login crew
list
search sailor
reference step
screengrab fyr
crew app
sailor
user
observe
see
note
section
text
card
booking management
incorrect step login
page expected
guest
getting
note refer
appear
one
change
shown
login sailorapp
need
please find
log shubhravya
show
screen recording
correct
day
select
option
order
field
attached screen
check
event
valid credential
sailor detail
currently
screen
space
instead
infinite loading
page click
shore excursion
dashboard page
build cert
launch sailor
verify
selected
iphone
description
existing account
error message

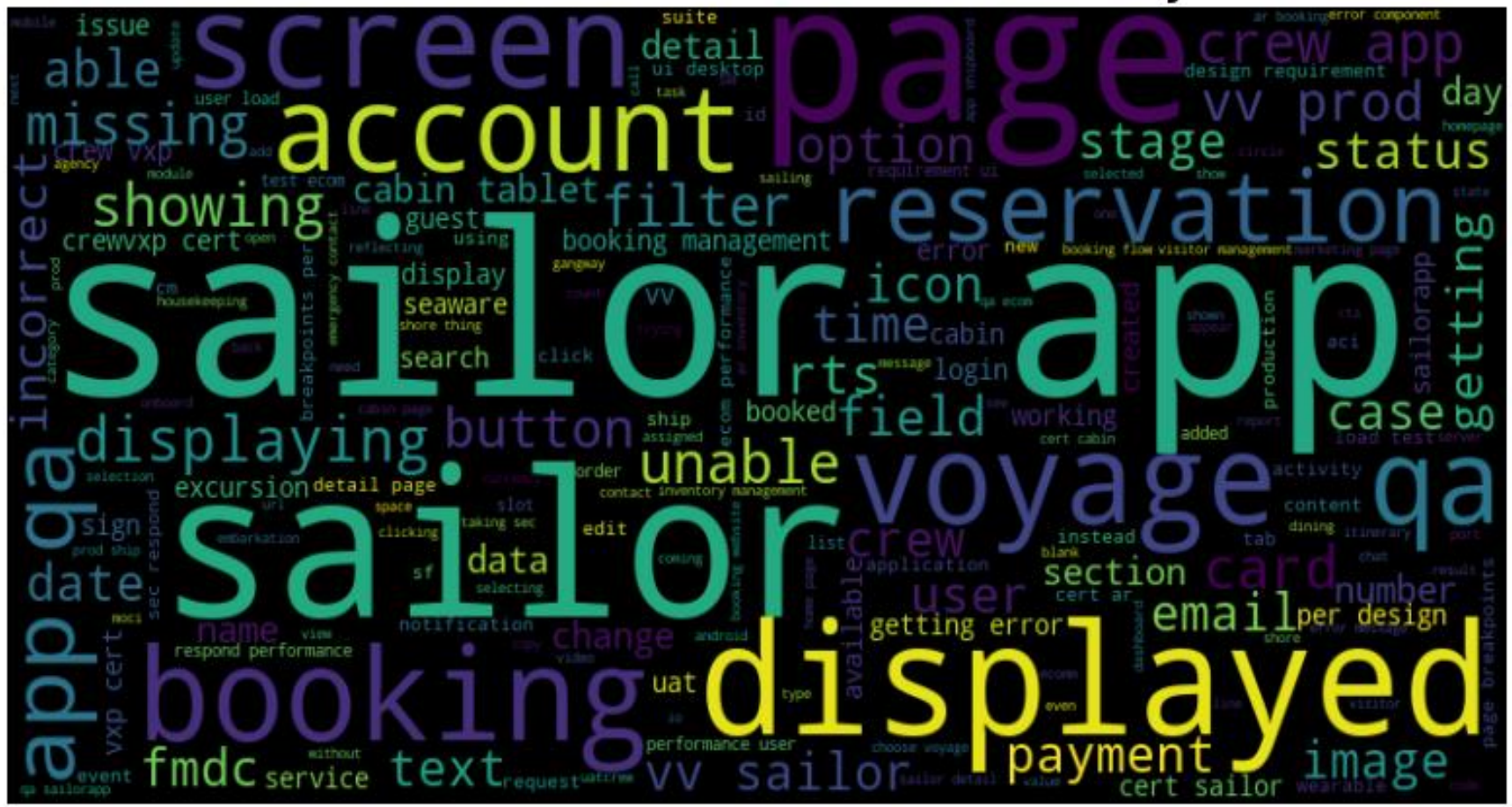
Word Cloud for InValid Descriptions



Word Cloud for Valid Summary



Word Cloud for Invalid Summary



Convert Text into TFI-IDF

Combined *Description* and *Summary* features into single text and performed TF-IDF Vectors

Logistic Regression

Confusion Matrix

[[155 281]
 [30 1068]]

Accuracy: 0.80
Classification Report:

	precision	recall	f1-score	support
0	0.84	0.36	0.50	436
1	0.79	0.97	0.87	1098
accuracy			0.80	1534
macro avg	0.81	0.66	0.69	1534
weighted avg	0.80	0.80	0.77	1534

Finding the possible reasons for invalid defects

Once defect classified as invalid,

1.Calculated cosine similarity between the TF-IDF vector of the input defect description and the TF-IDF vectors of all invalid defect descriptions.

2.Based on cosine similarity score , sorted most similar defects and extracted as possible reasons

Deployment –Streamlit App