

# SAFARICOM TWEETS CLASSIFICATION

# TEAM ALPHA PRESENTATION



**Antony Brian**

GROUP LEADER

Deployment



**Julia Karanja**

GROUP MEMBER

Presentation Slides



**Faith Makokha**

GROUP MEMBER

Jupyter Notebook/GitHub



**Arnold Kalage**

GROUP MEMBER

Write-Up



**Betty Bett**

GROUP MEMBER

Readme

# TABLE OF CONTENTS

Business  
Understanding

1

Data Preparation

3

Conclusion

5

Data Understanding

2

Modeling

4

Recommendation

6

# OVERVIEW

The advancement of technology has enhanced communication abilities with the use of social media in business and other sectors. Most companies have therefore resolved to creating accounts that specifically deal with customer care queries. Attending to these queries promptly is therefore at the core of most successful businesses as developing good customer interaction tools is key to increasing customer satisfaction.



# PROBLEM STATEMENT

In recent years, a number of companies have migrated to using social media for customer service and customer support, with companies having specific accounts dedicated to customer care queries. This has led to the problem of classifying said complaints or services according to the department that needs to work on them and generate tickets for the services.

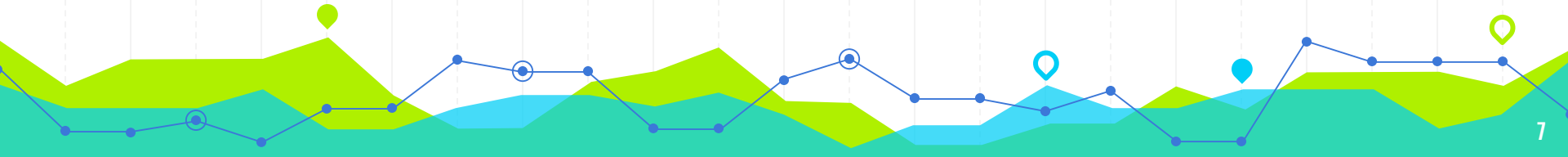


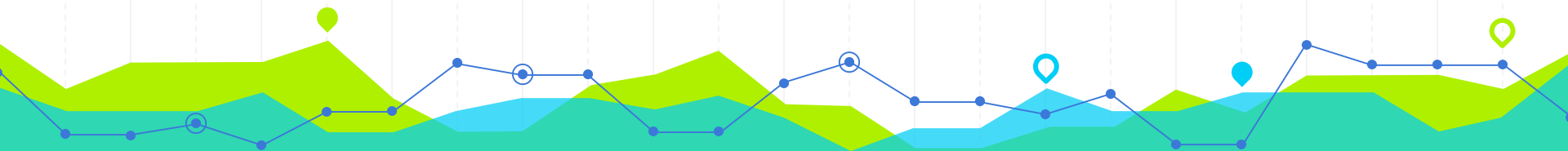
# SPECIFIC OBJECTIVES

- To identify the departments in which the complaints fall in
- To improve customer service efficiency
- To improve branding and customer relations
- To accelerate customer target marketing

“

*For the project success criteria we will be aiming for an F1 score of 70% to determine the efficiency and usability of our model.*





# DATA UNDERSTANDING

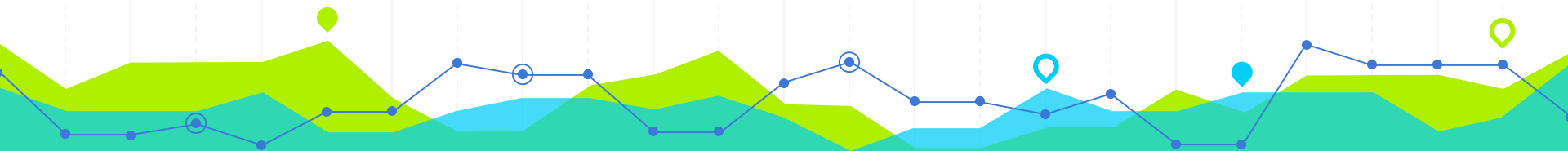




# 1 Million Tweets!

# DATA DESCRIPTION

Column	Description
Datetime	This refers to the date of the tweet
Tweet Id	This refers to the unique id of the tweet
Text	This refers to the actual text in the tweet
Username	This refers to the name of the user who commented
Like Count	This refers to the number of likes a tweet receives



# DATA PREPARATION

# DATA CLEANING

## @SAFARICOM\_CARE

Removing all the tweets that are from Safaricom\_Care.

S

R

## REMOVING NOISE

Removing: links, usernames, hashtags, punctuations, emojis, stopwords and numbers

T

L

Time of day the tweet was sent, either morning, afternoon or evening.

## TIME\_OF\_DAY

Creating labels for our data using key words manually

## LABELS

# EXPLORATORY DATA ANALYSIS

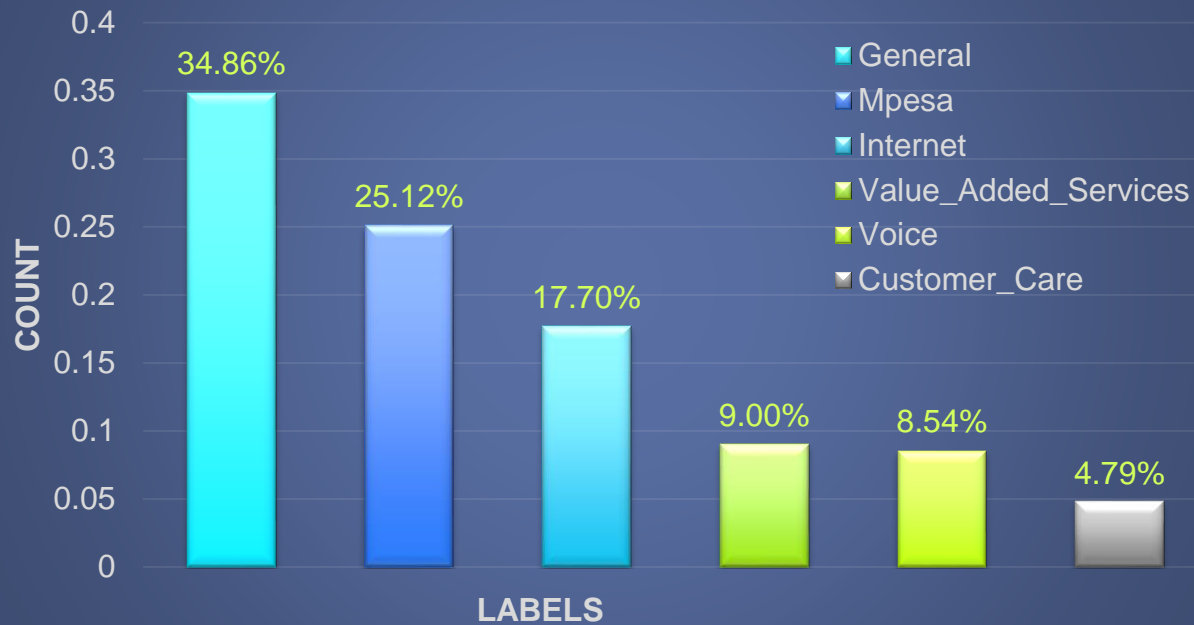


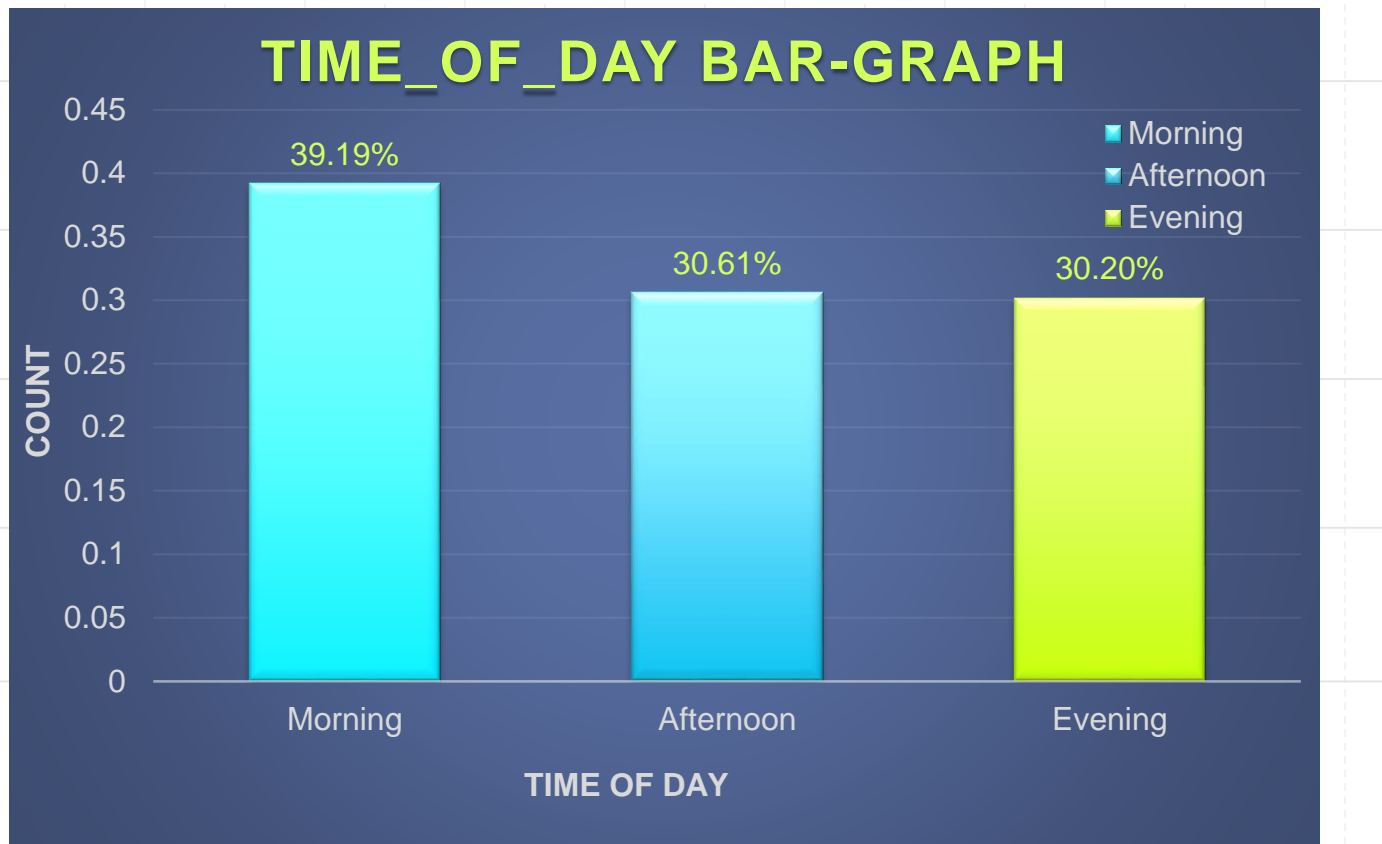
```
graph LR; A[Univariate Analysis] --> B[Bivariate Analysis]
```

Univariate  
Analysis

Bivariate  
Analysis

## CATEGORY BAR-GRAPH

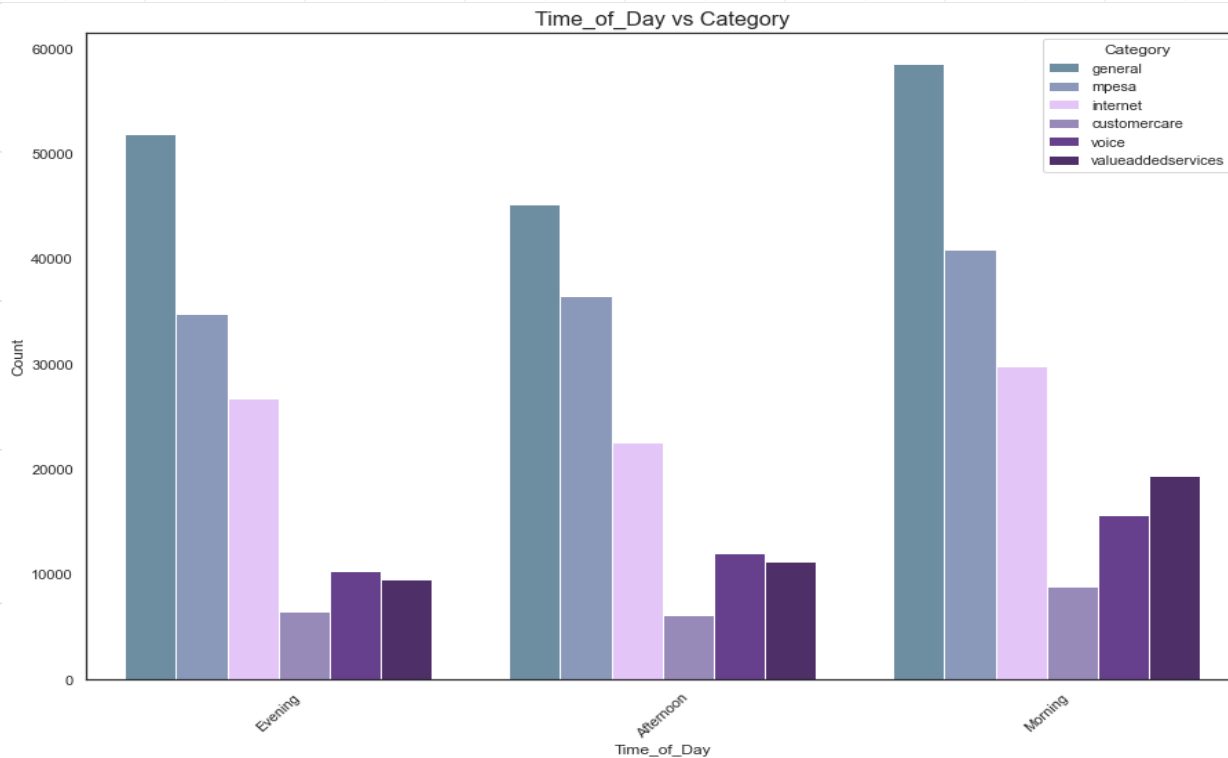








# TIME\_OF\_DAY VS CATEGORY



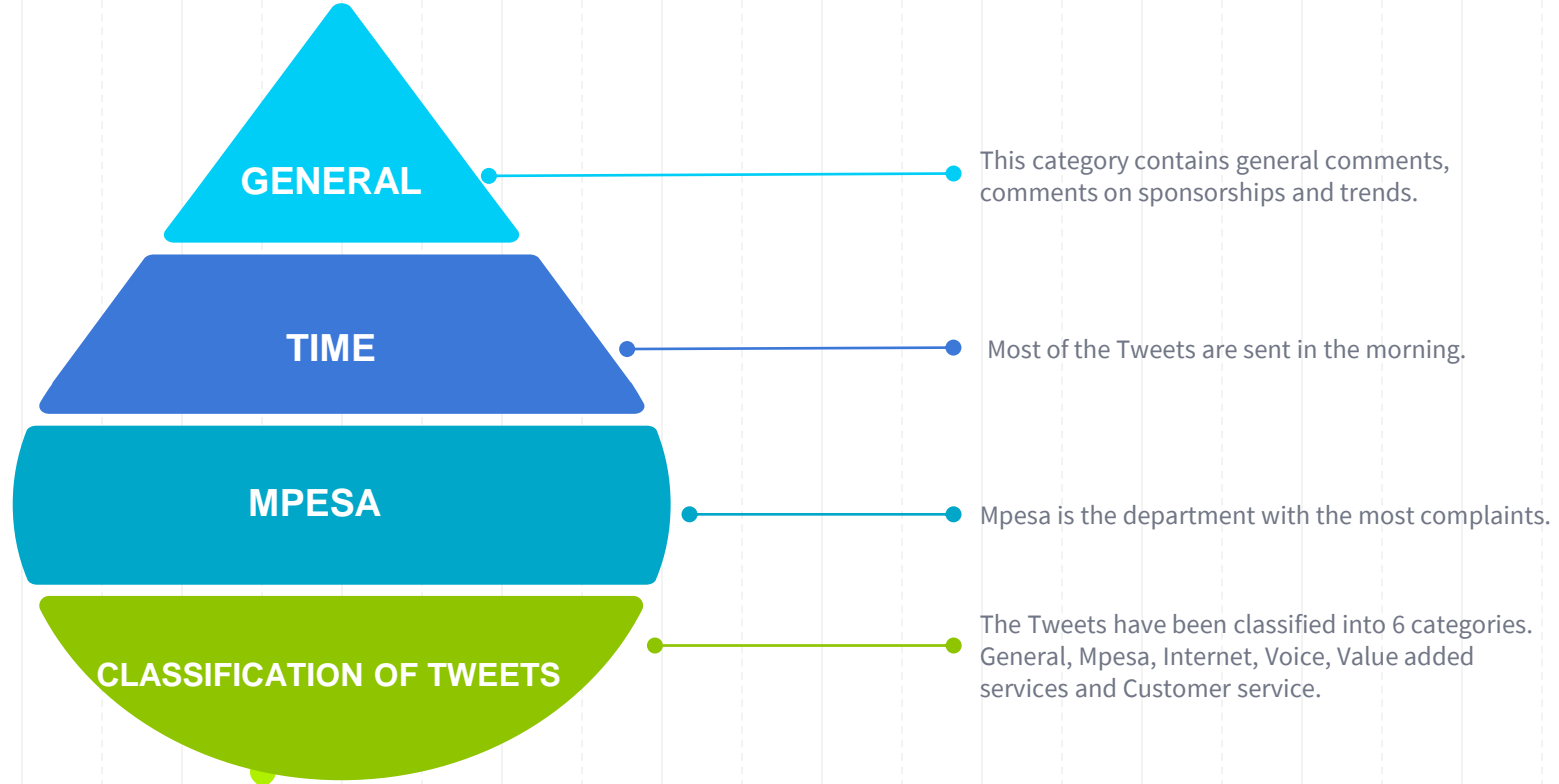


# MODELING AND EVALUATION

# MODEL EVALUATION

	TRAIN ACCURACY	TEST ACCURACY	RECALL	PRECISION	F1 SCORE
KNN	0.568	0.547	0.547	0.724	0.510
DECISION TREES	0.711	0.710	0.710	0.775	0.690
RANDOM FOREST	0.931	0.808	0.808	0.829	0.806
XGBOOST	0.947	0.947	0.947	0.949	0.947
NEURAL NETWORKS	0.999	0.999	0.999	0.999	0.999

# CONCLUSION



# RECOMMENDATION

## Recommendations

Classify the tweets according to the respective departments

More attention should be given to the MPESA department .

Track time for planning employee shifts and strategizing marketing hours.

Look into the general category due to sponsorships and other general comments.

## Benefits

Improved Customer Efficiency and better customer targeting in Marketing

Improved efficiency; New customers will increase their Revenue.

Reduce employee idle time, reduce cost and improved marketing strategy.

The Brand is maintained.

# THANK YOU!

**Any questions?**

