

# NLP AND CLASSIFICATION PROJECT

## FOCUS AREA: ANALYSING NEWS ARTICLES DATASET

" ONLY THE HUMAN MIND INVENTS CATEGORIES AND TRIES TO FORCE FACTS INTO SEPARATED PIGEONHOLES. – ALFRED KINSEY



COMPILED BY: KELLY NCUBE, NEO MOUMAKOA,  
LISA KWANINI, MFUNDO SINDANE

TEAM NUMBER: 14

# summary

The focus area of this project is to create classification models using Python and deploy it as a web application with Streamlit. The main objective is to design a competitive yet “best-fit” model to categorise newspaper articles. The news classification project involves these key stages: data loading, preprocessing, model training, evaluation, and final deployment. The primary stakeholders are the editorial team, IT/tech support, management, and readers.

# POTENTIAL CHALLENGES

THESE COULD AFFECT THE PROJECT'S GOAL ATTAINMENT DUE TO MULTINOMIAL DATA:

- Data imbalance
- Scaling numerical data
- Textual feature dimensionality
- Feature heterogeneity



# Data preprocessing

PHASES:



DATA LOADING :

- 
- 

DATA PREPROCESSING:

- 
- 

**"Data preprocessing is a critical step in data analysis because it ensures the quality, reliability, and usability of the data"**

# MODEL TRAINING

This project trained 3 models. These are the chosen models:

1 MODEL 1[ ]

The

2 MODEL 2 [ ]

The

3 MODEL 3 [ ]

The

The best performing model

# CONCLUSION

The models tested were..  
the best performing model was..  
Summing up..

