

REVOLUTIONIZING NEWSROOMS WITH NLP AND MACHINE LEARNING

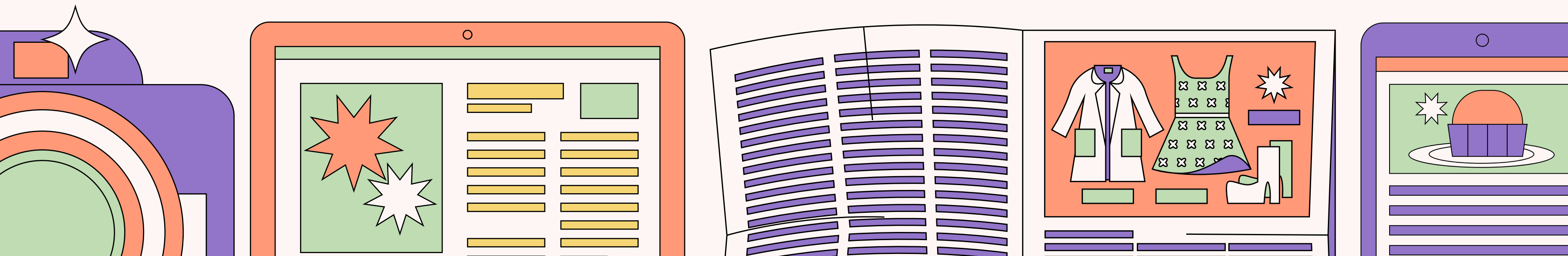
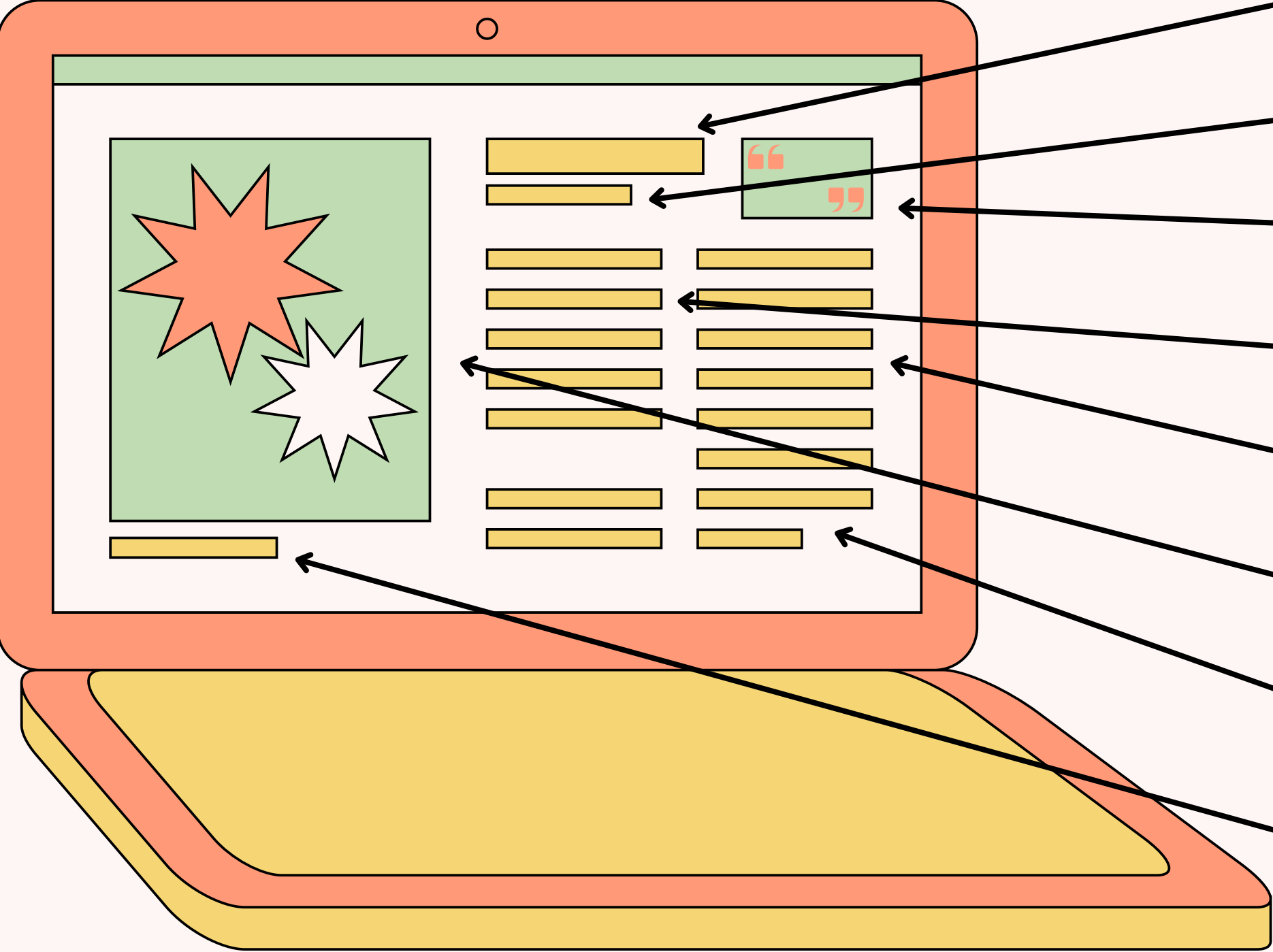


Table of Contents



INTRODUCTION

PROJECT OVERVIEW

INSIGHTS

MODEL BUILDING

MODEL IMPROVEMENT

MODEL DEPLOYMENT

RECOMMENDATIONS

CONCLUSION

MEET OUR TEAM

- **CLEMENT MPHETHI - LEAD DATA SCIENTIST**
- **NOLWAZI MNDEBELE - PROJECT MANAGER**
- **TSHEPISO MUDAU - GITHUB MANAGER**
- **NEO RADEBE - DATA SCIENTIST**
- **NALEDI MOGAFA MOGALE - DATA SCIENTIST**
- **KOENA MCDONALD MAHLADISA - DATA SCIENTIST**

INTRODUCTION



**HOW CAN WE HARNESS THE POWER OF DATA SCIENCE
TO ENHANCE THE READER EXPERIENCE AND STREAMLINE
NEWS CONTENT MANAGEMENT?**

PROJECT OVERVIEW



**HOW CAN WE DEVELOP AN ACCURATE MODEL FOR BETTER
NEWS CATEGORIZATION?**

PROJECT WALKTHROUGH

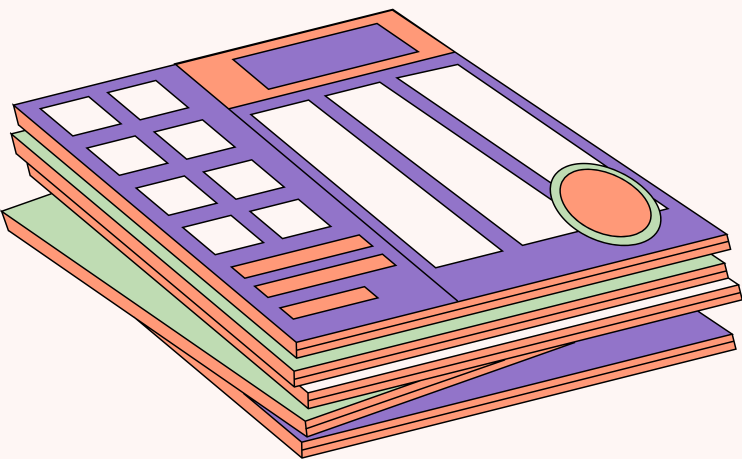
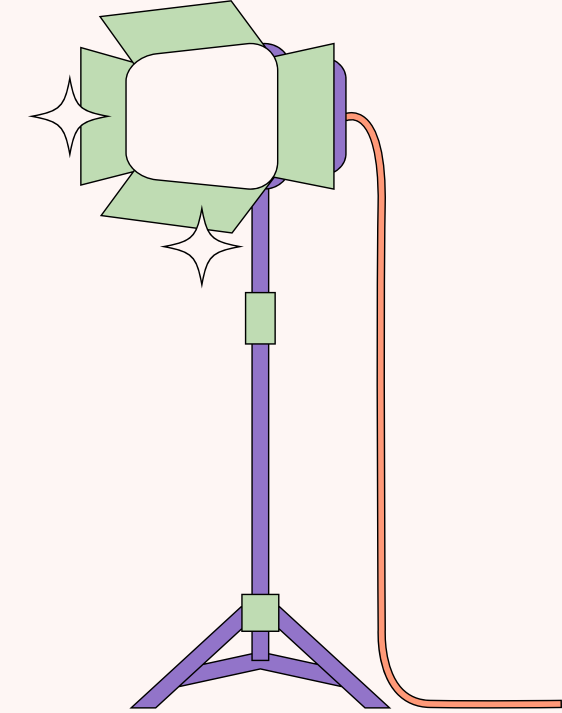
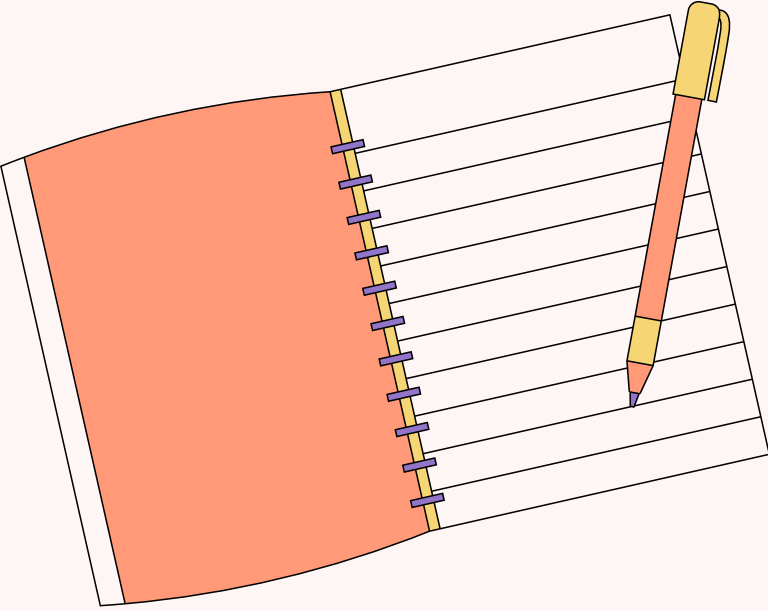
**DATA
CLEANING**

**EXTRACT
INSIGHTS**

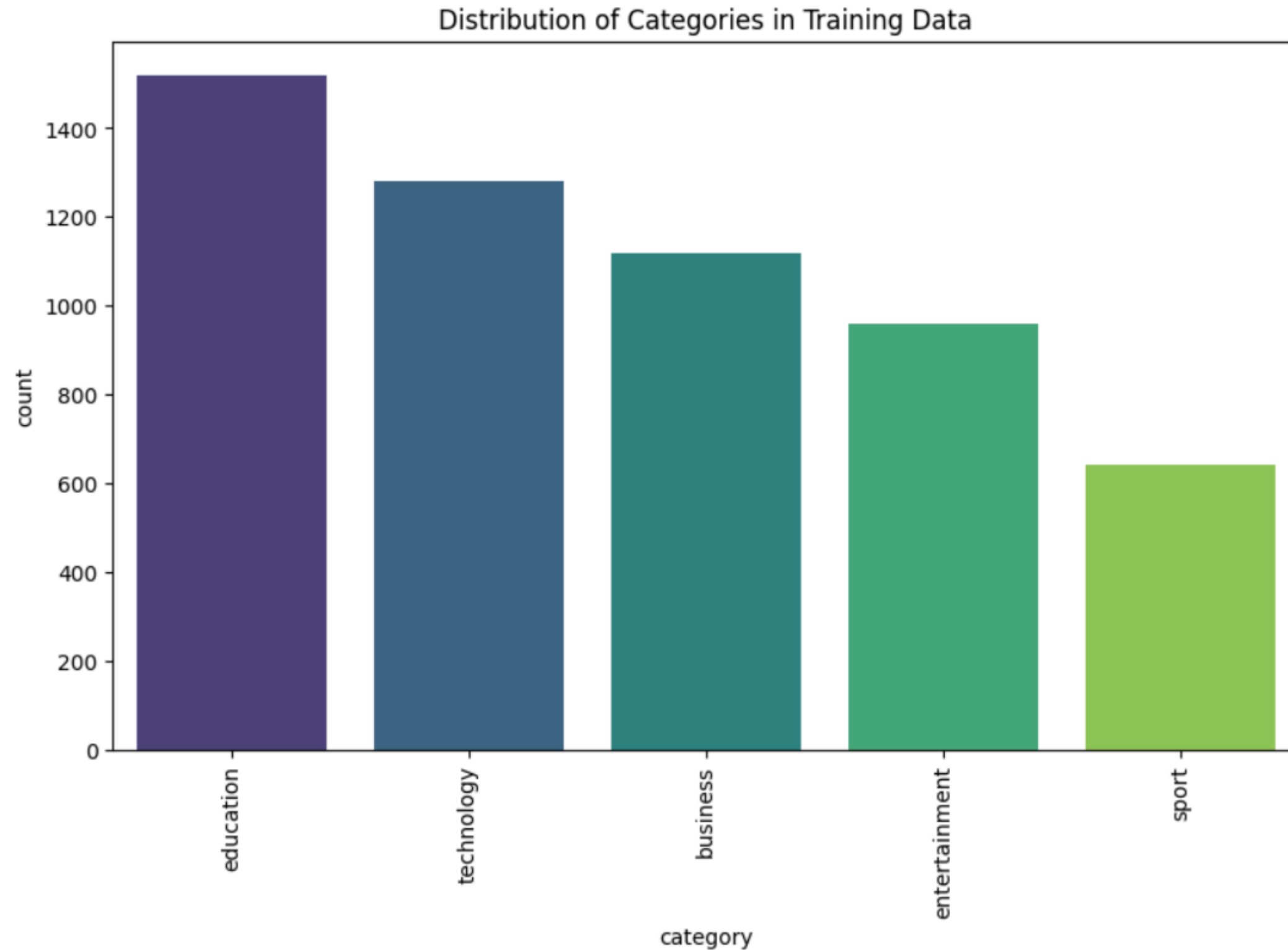
**PREDICTIVE
MODELS**

**MODEL
IMPROVEMENT**

**MODEL
DEPLOMENT**



INSIGHTS

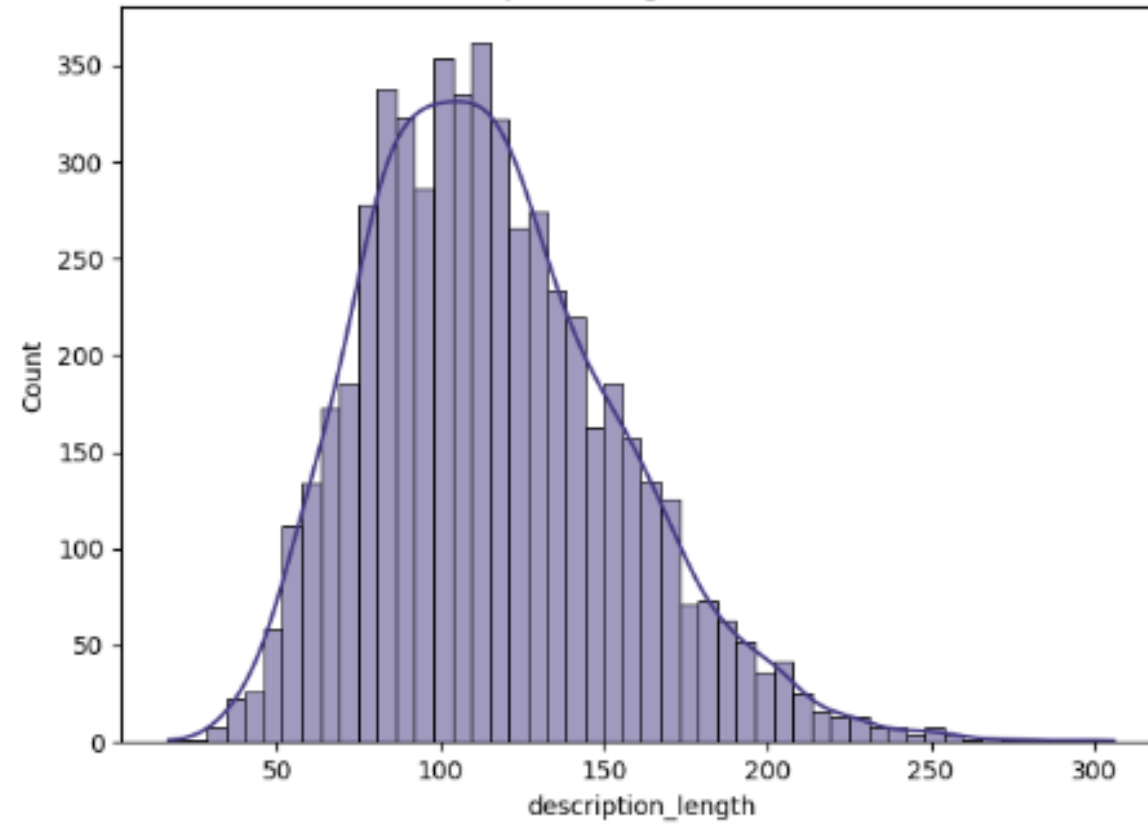


- **EDUCATION IS THE TOP CATEGORY, WITH MORE THAN 1400 ENTRIES.**
- **SPORT IS THE CATEGORY WITH THE LEAST ENTRIES, LESS THAN 800.**

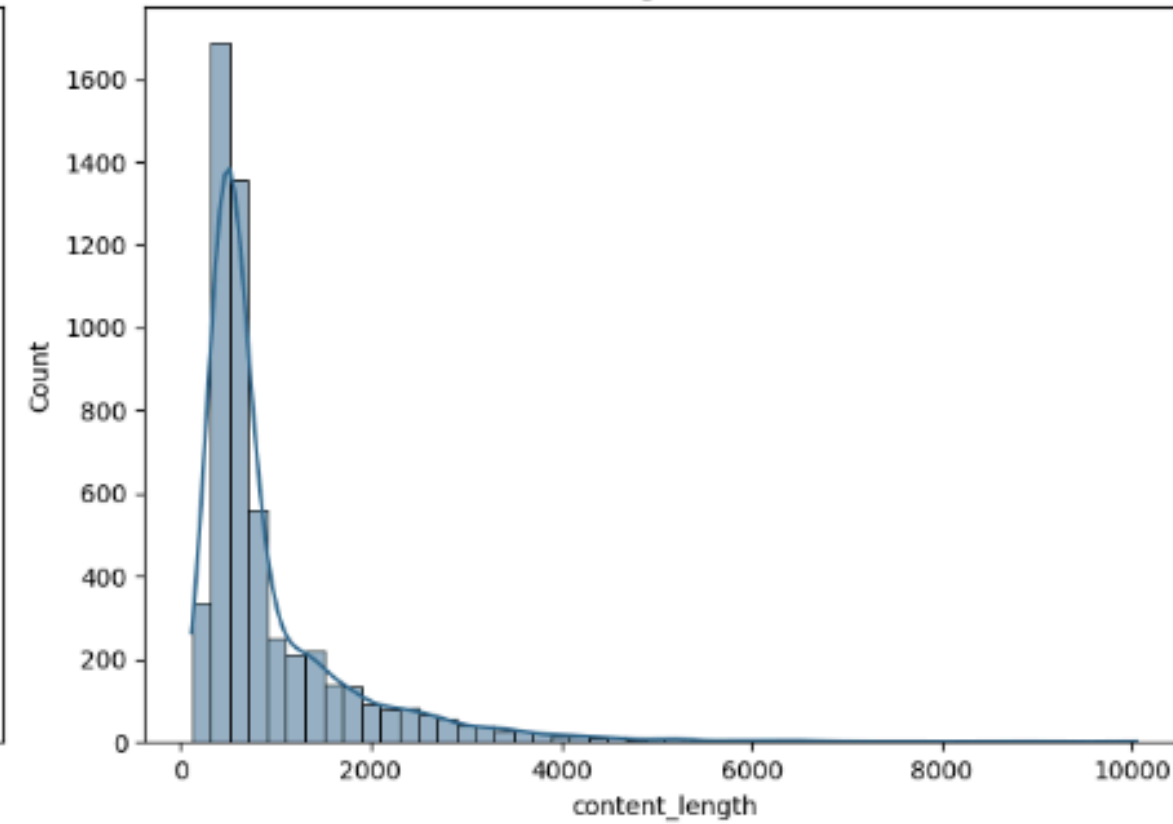


TEXT LENGTH ANALYSIS

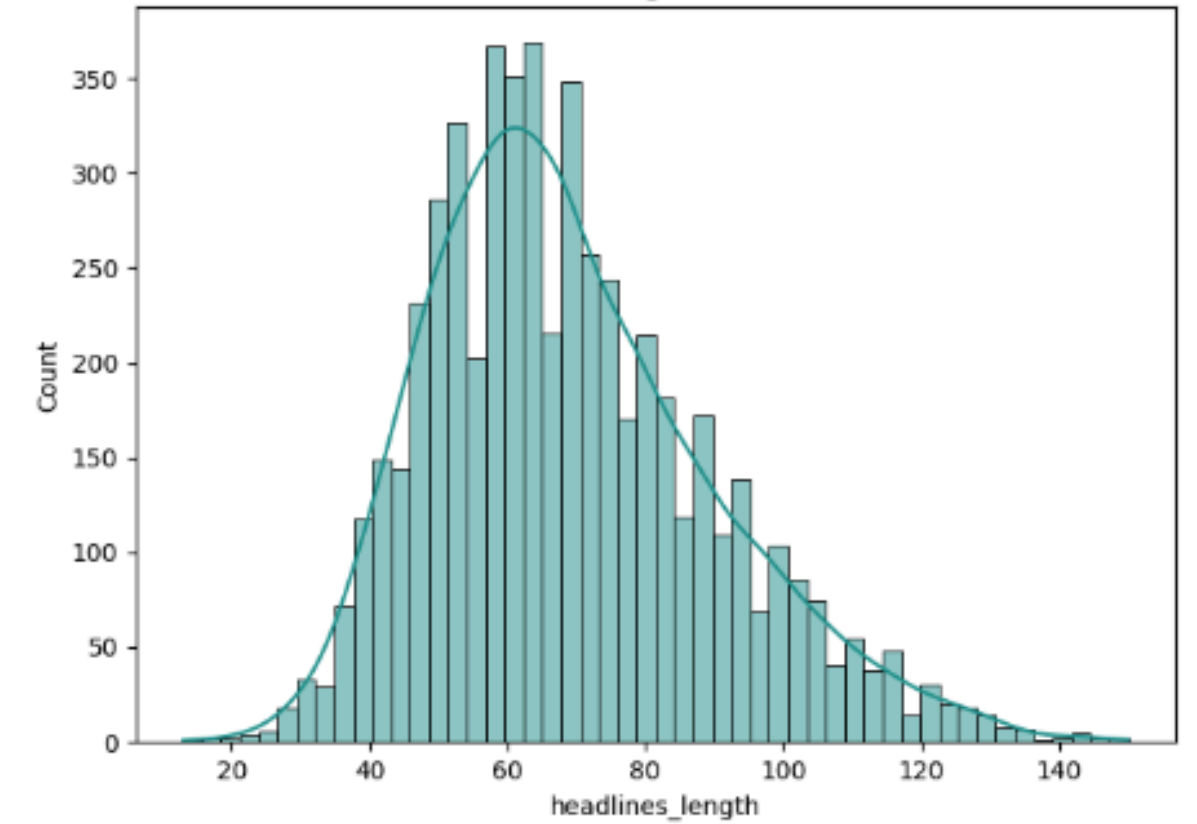
Description Length Distribution



Content Length Distribution



Headlines Length Distribution



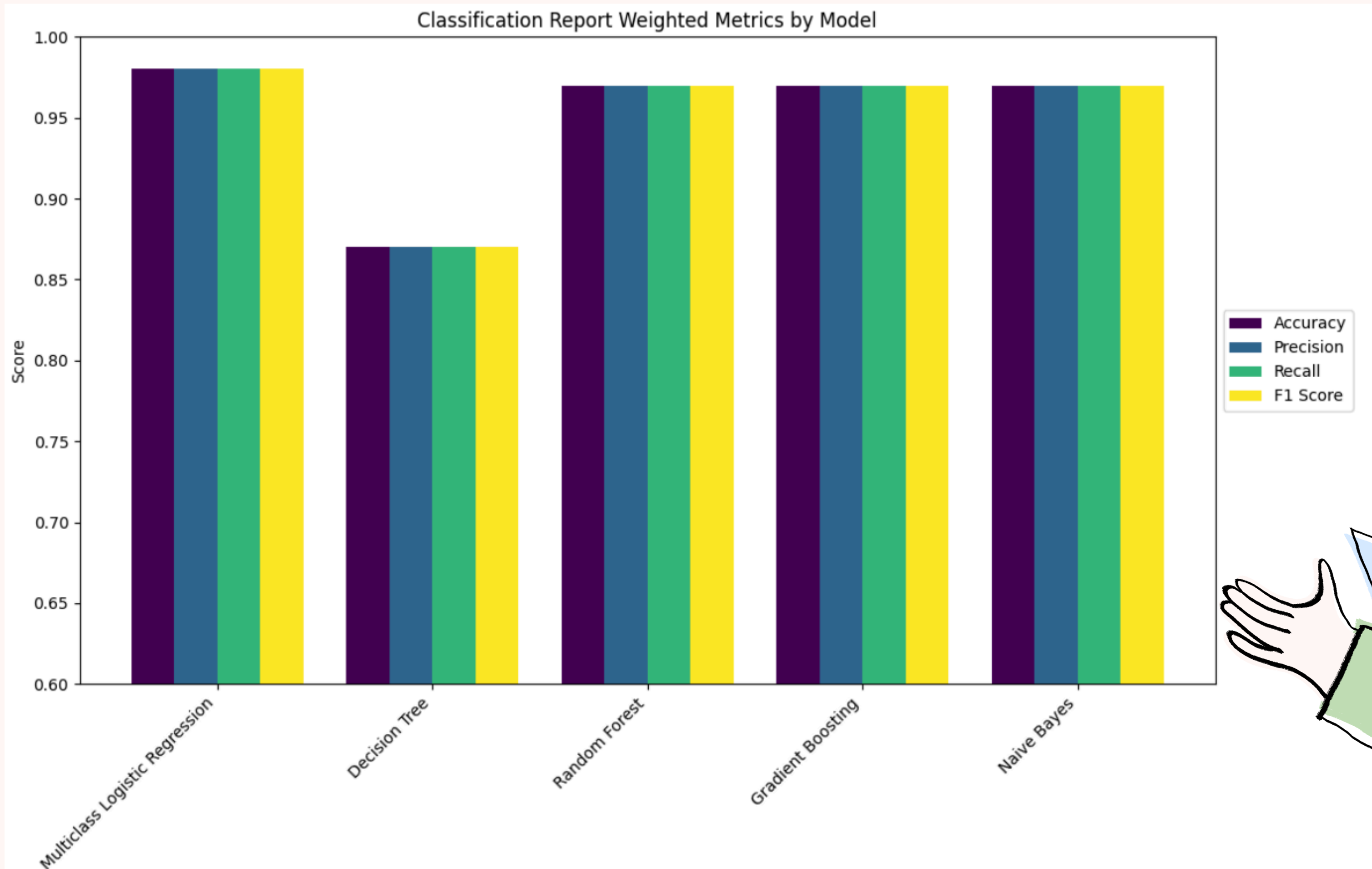
WORD CLOUD VISUALIZATION



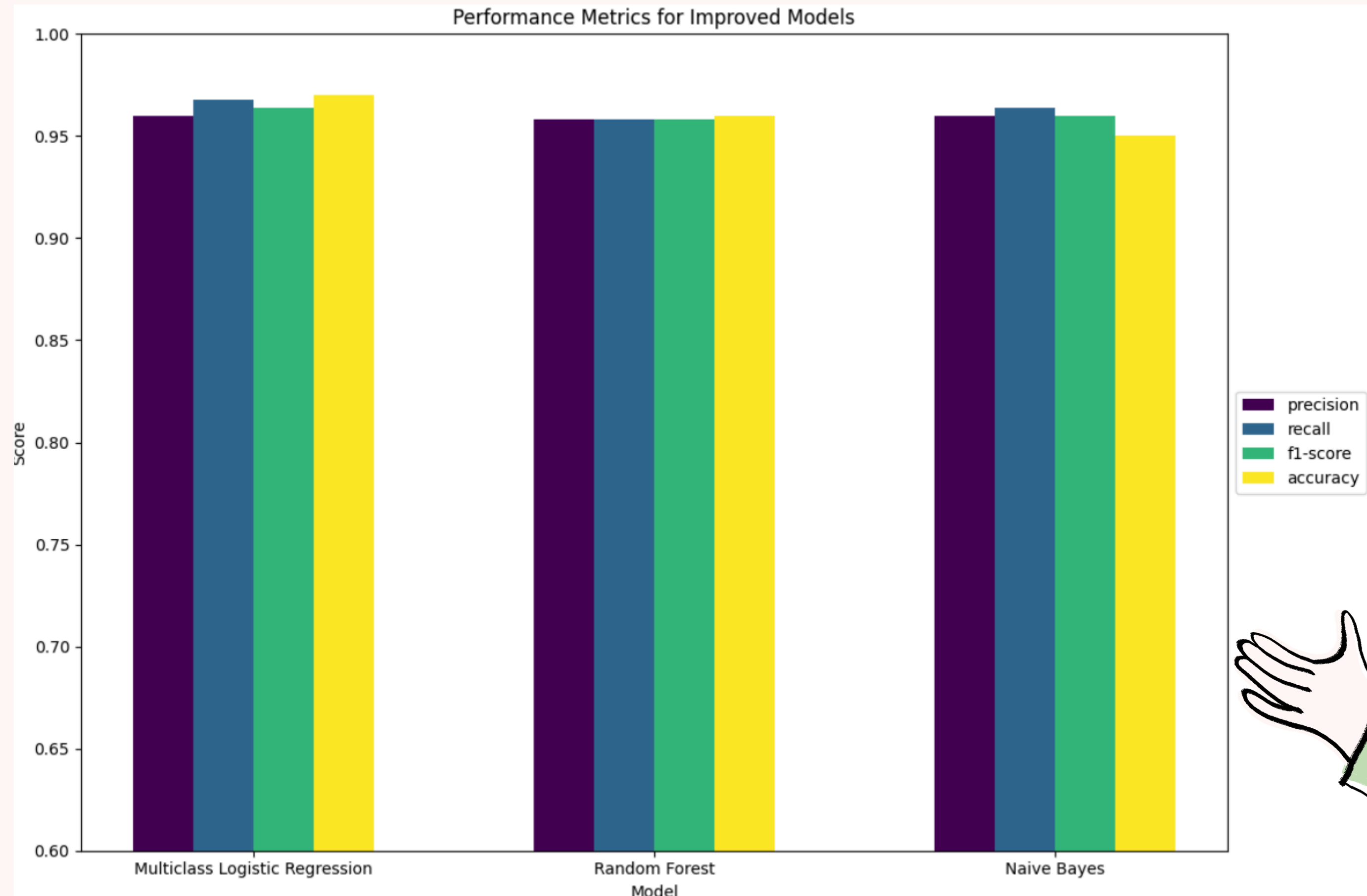
MODEL TRAINING

- MULTICLASS LOGISTIC REGRESSION
- DECISION TREES
- RANDOM FOREST MODEL
- GRADIENT BOOSTING MODEL
- NAIVE BAYES MODEL

MODEL INSIGHTS



IMPROVED MODELS



MODEL DEPLOYMENT



[HTTPS://HOODPHARMXCIST-CLASSIFICATION-APP-BASE-APP-TODKR5.STREAMLIT.APP/](https://hoodpharmxcist-classification-app-base-app-todkr5.streamlit.app/)

RECOMMENDATIONS

BASED ON THE PROJECT FINDINGS, HERE ARE SOME OUR RECOMMENDATIONS FOR FURTHER IMPROVEMENT AND FUTURE WORK:

- USE ADVANCED TECHNIQUES LIKE PHRASE EXTRACTION FOR NUANCED INFORMATION**
- CONTINUOUSLY MONITOR AND UPDATE THE MODEL POST-DEPLOYMENT TO ADAPT TO CHANGING TRENDS.**
- GATHER AND INCORPORATE USER FEEDBACK TO IMPROVE THE SYSTEM BASED ON REAL-WORLD USAGE**

CONCLUSION

THE NEWS CLASSIFIER APP HAS DEMONSTRATED EXCEPTIONAL SUCCESS WITH A 97% ACCURACY RATE AND ROBUST PRECISION AND RECALL SCORES EXCEEDING 0.95 ACROSS ALL CATEGORIES.

ITS PRACTICAL APPLICATIONS IN JOURNALISM STREAMLINE OPERATIONS, IMPROVE CONTENT RELEVANCE, AND BOOST AD REVENUE. TECHNICALLY, INTEGRATING ADVANCED MODELS LIKE MULTICLASS LOGISTIC REGRESSION ENSURES SCALABILITY AND RELIABILITY EVEN UNDER HIGH LOAD CONDITIONS. FUTURE PLANS INCLUDE ADDING MULTILINGUAL SUPPORT AND REAL-TIME CLASSIFICATION, DRIVEN BY CONTINUOUS USER FEEDBACK FOR ONGOING ENHANCEMENT.

OVERALL, THE APP SURPASSES EXPECTATIONS, ESTABLISHING ITSELF AS A PIVOTAL TOOL IN THE FIELD OF NEWS CLASSIFICATION.