



Natural Language Processing (NLP) Challenge

GROUP 3

by: Aryam - Dana - Randa - Ghadh

Project Overview

Tasks

- Process and clean the data.
- classification model.
- Results model.
- Predicting classification with new data

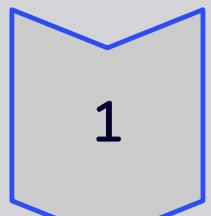
Data

- A classified dataset data.csv includes a label (0: fake, 1: real).
- The unclassified dataset validation_data.csv a includes label (2).

Goal

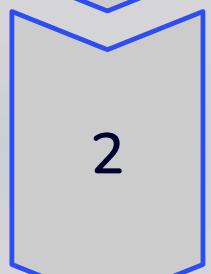
- To build a high-performing classifier model for distinguishing between fake and real news and test it on validation data.

Data Preprocessing using NLP Techniques



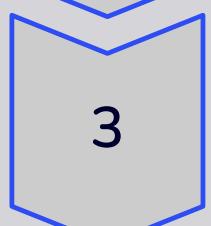
1 Download Data

Download data and learn subject data



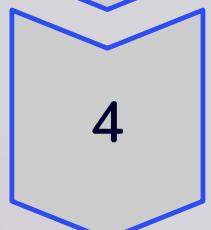
2 Clean Data

- Removing symbols.
- Converting uppercase letters to lowercase.



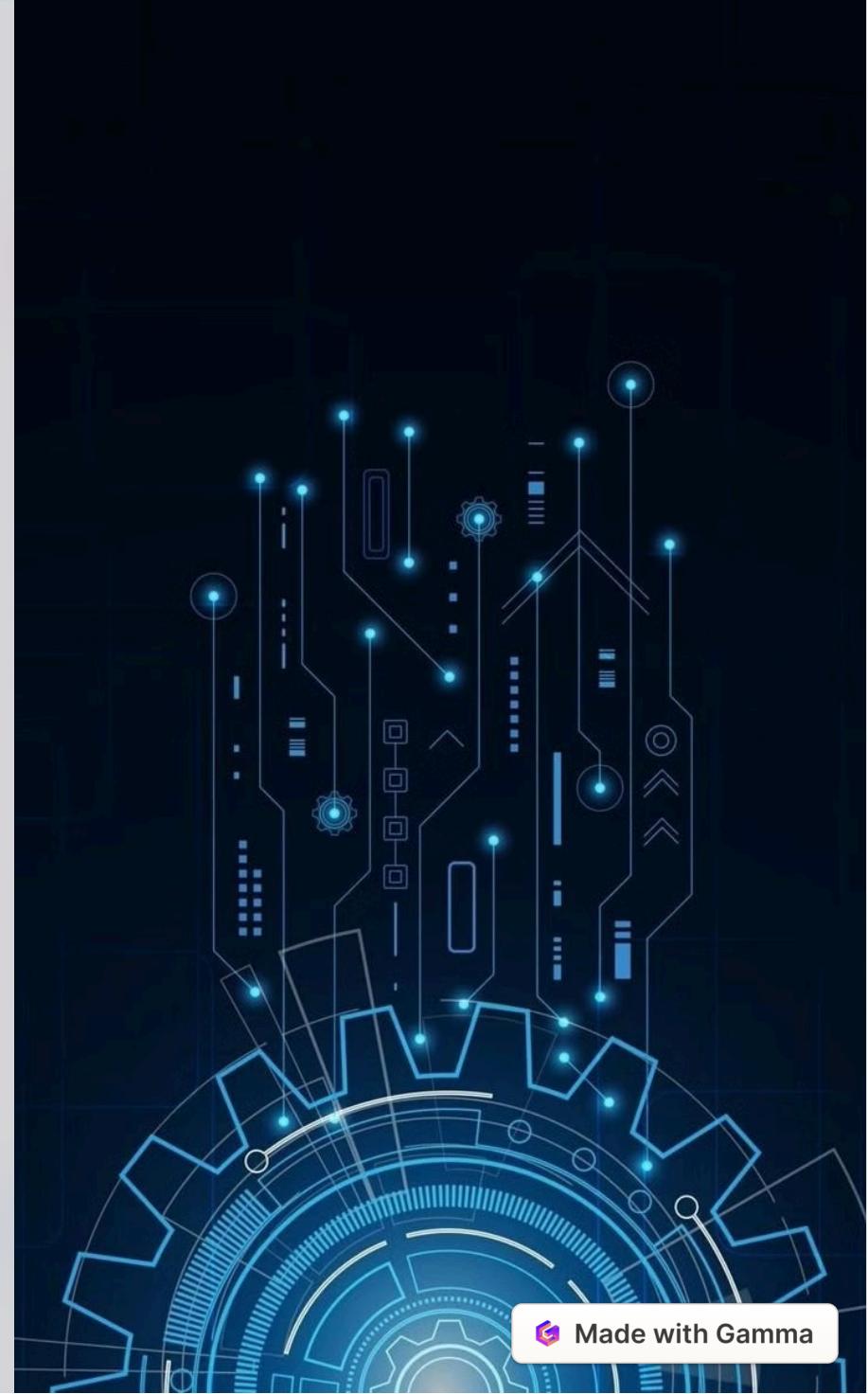
3 Extract Features

Convert text to numbers using TF-IDF



4 Train Data

Split data into training and testing.





Classification Models

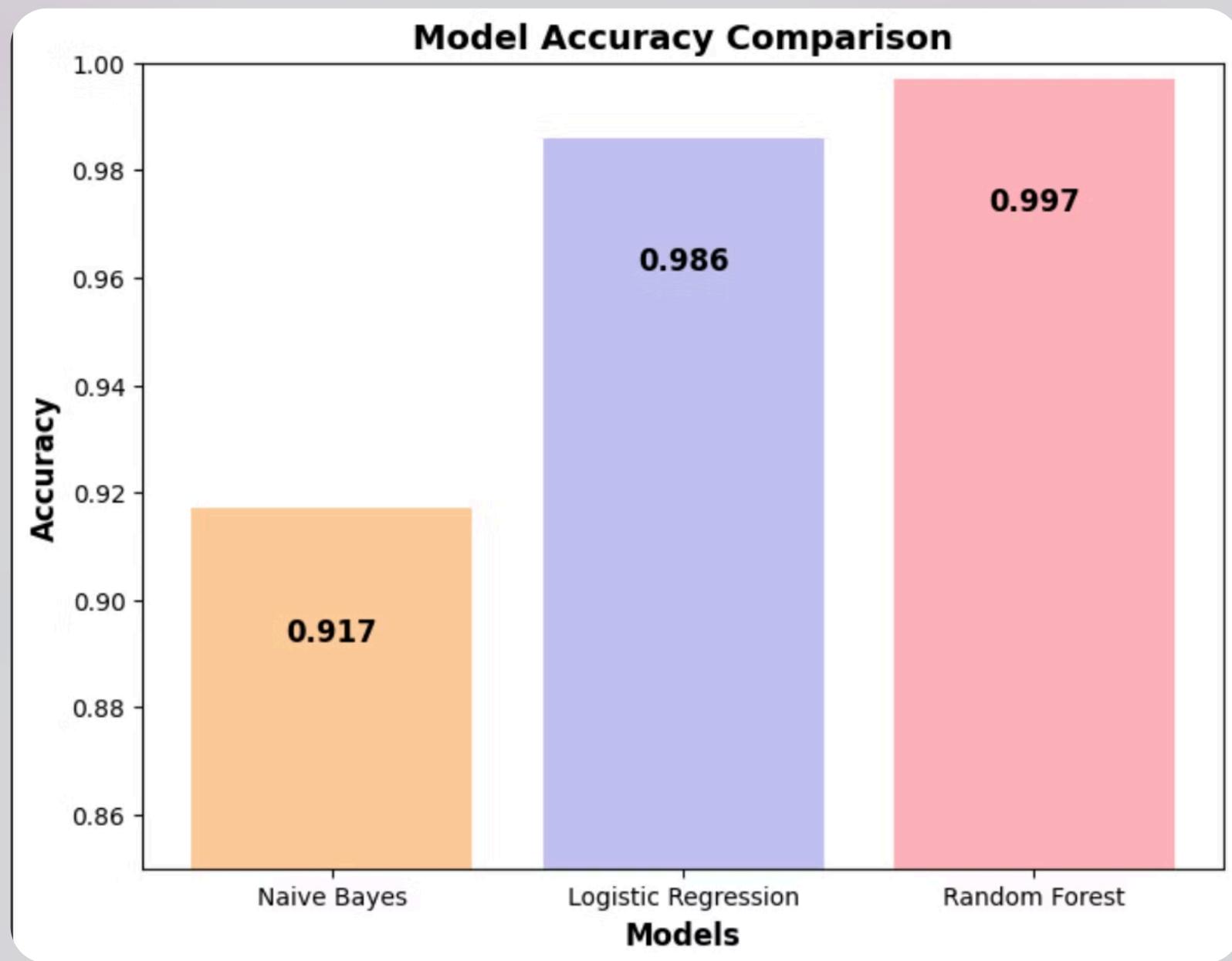
Tested three different models for news headline classification:

1- Logistic Regression

2- Naive Bayes

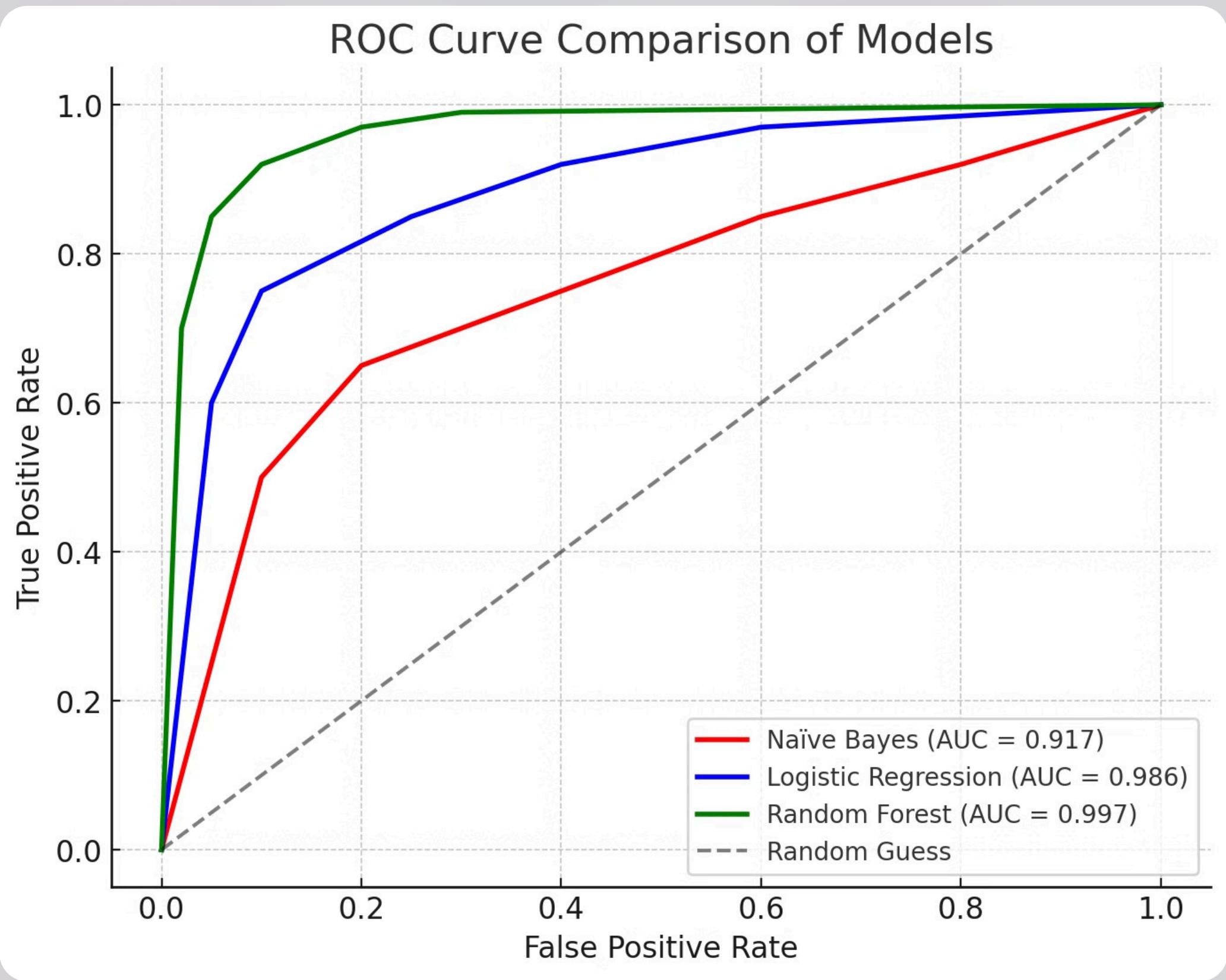
3- RandomForest

Observation Between Models:



Precision	Recall	F1- Score	Models
99.5%	99.7%	99.6%	RandomForest
91%	91%	91.2%	Naive-Bayers
98%	98%	98.5%	Logistic regression

Our Model Final Evaluation – The ROC - AUC Curve





The Best Selected Classification - RandomForest Model

99.6%

Accuracy

99.7%

Recall

99.5%

Precision

99.6%

F1-Score

Predicting Classification:

Dataset before development:

Google Docs +

validation_data

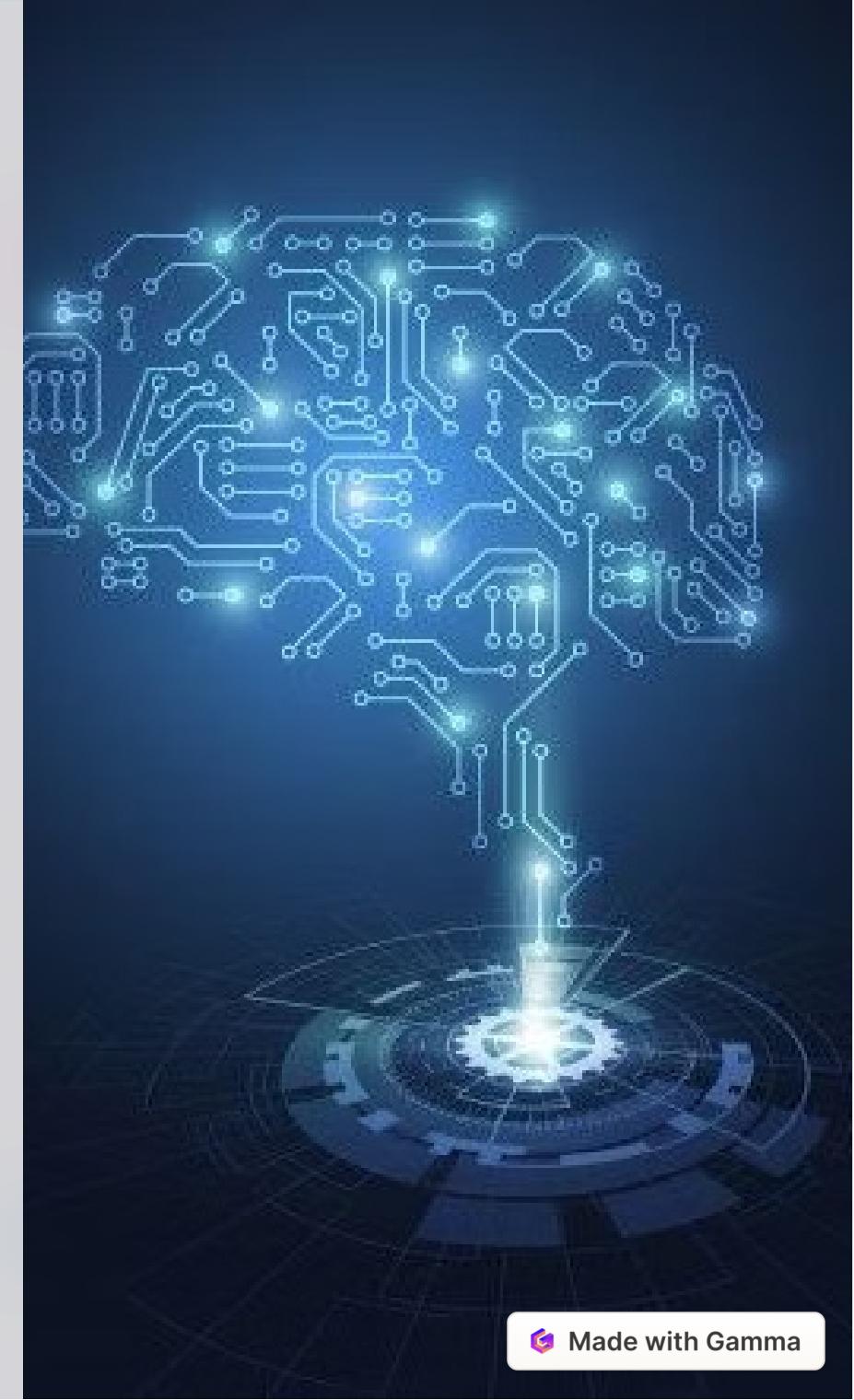
title	text	subject	date
2	UK's May 'receiv	LONDON (Reute worldnews	September 15, 2017
2	UK transport poli	LONDON (Reute worldnews	September 15, 2017
2	Pacific nations cr	WELLINGTON (F worldnews	September 15, 2017
2	Three suspected	ADEN, Yemen (F worldnews	September 15, 2017
2	Chinese academ	BEIJING (Reuter worldnews	September 15, 2017
2	Flames raced alc	LONDON (Reute worldnews	September 15, 2017
2	London police ad	LONDON (Reute worldnews	September 15, 2017
2	London ambulan	LONDON (Reute worldnews	September 15, 2017
2	Witness says inj	LONDON (Reute worldnews	September 15, 2017
2	UK says world wi	LONDON (Reute worldnews	September 15, 2017
2	China urges peal	BEIJING (Reuter worldnews	September 15, 2017
2	Swedish PM sun	STOCKHOLM (F worldnews	September 15, 2017
2	Russia's Lavrov i	MOSCOW (Reut worldnews	September 15, 2017
2	Peru's Congress LIMA	(Reuters) - worldnews	September 15, 2017
2	Taiwan jails main	TAIPEI (Reuters) worldnews	September 15, 2017
2	Philippines' Preside	n MANILA (Reuter worldnews	September 15, 2017

Dataset after development:

Google Docs +

validation_data_predicted

title	text	subject	date
1	UK's May 'receiv	LONDON (Reute worldnews	September 15, 2017
1	UK transport poli	LONDON (Reute worldnews	September 15, 2017
1	Pacific nations cr	WELLINGTON (F worldnews	September 15, 2017
1	Three suspected	ADEN, Yemen (F worldnews	September 15, 2017
1	Chinese academ	BEIJING (Reuter worldnews	September 15, 2017
1	Flames raced alc	LONDON (Reute worldnews	September 15, 2017
1	London police ad	LONDON (Reute worldnews	September 15, 2017
1	London ambulan	LONDON (Reute worldnews	September 15, 2017
1	Witness says inj	LONDON (Reute worldnews	September 15, 2017
1	UK says world wi	LONDON (Reute worldnews	September 15, 2017
1	China urges peal	BEIJING (Reuter worldnews	September 15, 2017
1	Swedish PM sun	STOCKHOLM (F worldnews	September 15, 2017
1	Russia's Lavrov i	MOSCOW (Reut worldnews	September 15, 2017
1	Peru's Congress LIMA	(Reuters) - worldnews	September 15, 2017
1	Taiwan jails main	TAIPEI (Reuters) worldnews	September 15, 2017
1	Philippines' Preside	n MANILA (Reuter worldnews	September 15, 2017



Team work

Name	Aryam	Dana	Randa	Gadah
Task	Codeing	Codeing	Presentation	Presentation
Percent	25%	25%	25%	25%

Thank You!