WB News Classifier

Semester : 5

Branch : CSE

Software Group Project - III

CS348

Group Members

19DCS088 Gracy Patel

19DCS118 Juhi Sanchla

19DCS158 Sakshi Zalavadia

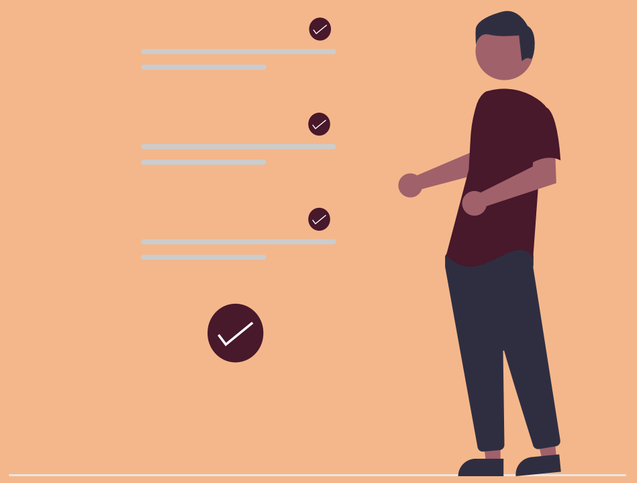
Guided By

Prof. Aishwariya Budhrani

News Genre Classification

WB News classifier is a web application that classifies news articles based on the particular genre with accuracy using Natural Language Processing.

Agenda



- Problem Statement
- Features
- Requirement Analysis
- Goals and Outcome
- Flow chart
- Technology Stack
- Resources

Problem Statement

News genre classifier is a web application that takes input news from the user and classifies news article based on the particular genre with accuracy using Natural Language Processing. It also provides with the recent news from all over the world

It helps identifying what kind of news article conveys by its content and images.

05

Features

News Genre Detection

Classifying the genre based on news article with accuracy

Fetch recent news and classifiy

Application fetchs recent news using web scraping and classifies it genre

Classification with Text input

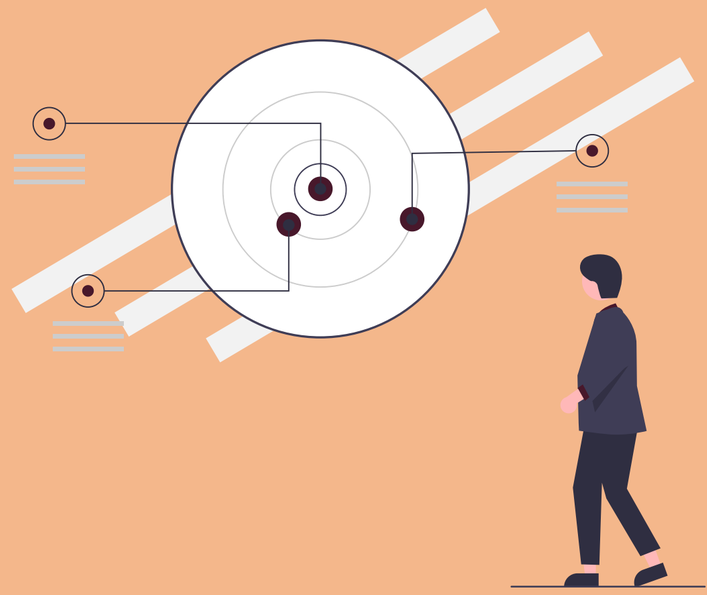
Predicting the group based on the input provided

API Request for classification

Creating REST APIS for the above to integrate into a web app and returning json object

06

Goals and Outcome



GOAL

Our main aim is to train a model which predicts the news genre based on the news provided and create an accessible web application.

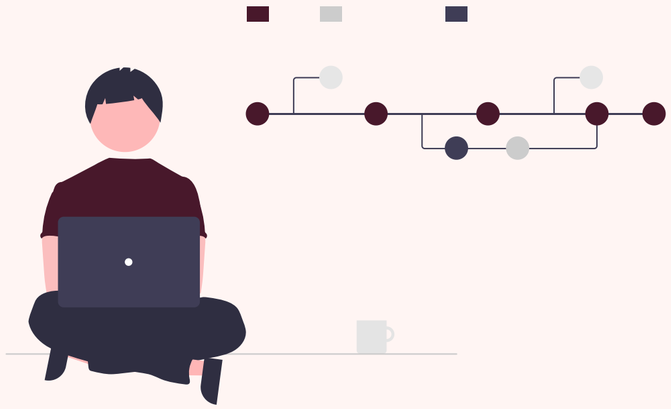
OUTCOMES

- To classify a news article into a particular group
- To keep track of accuracy
- To create a web app to be useable
- To scrap data for fetching recent news

Requirement Analysis

- Brainstorming
- Document Analysis
- Prototyping
- Workshops/Seminars and hands-on
- Reverse Engineering

Technology Stack



Language

- Python
- JavaScript

Designing

- Figma
- Material UI

Data Set

- BBC News --Kaggle

Web Scarping

- Beautiful Soap

IDEs

- Jupyter notebook
- VScode
- Postman

Version Control

- Github

Web Application

- ReactJS
- Redux

Authentication

- Firebase

Flask APIs

- Flask

Technology Stack

Classification methods used

- Naive Bayes Classification
- Support Vector Machine

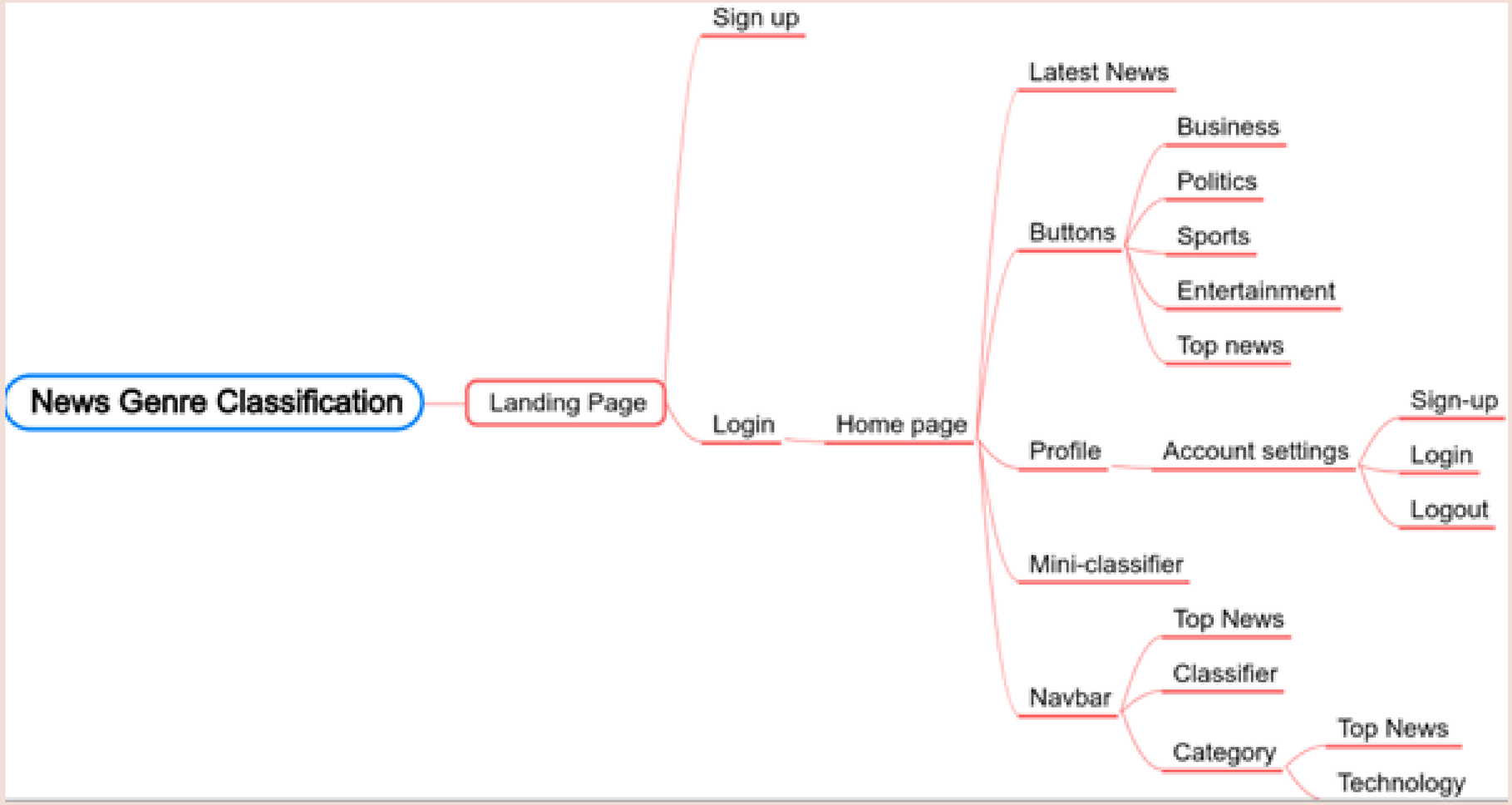
Approaches used

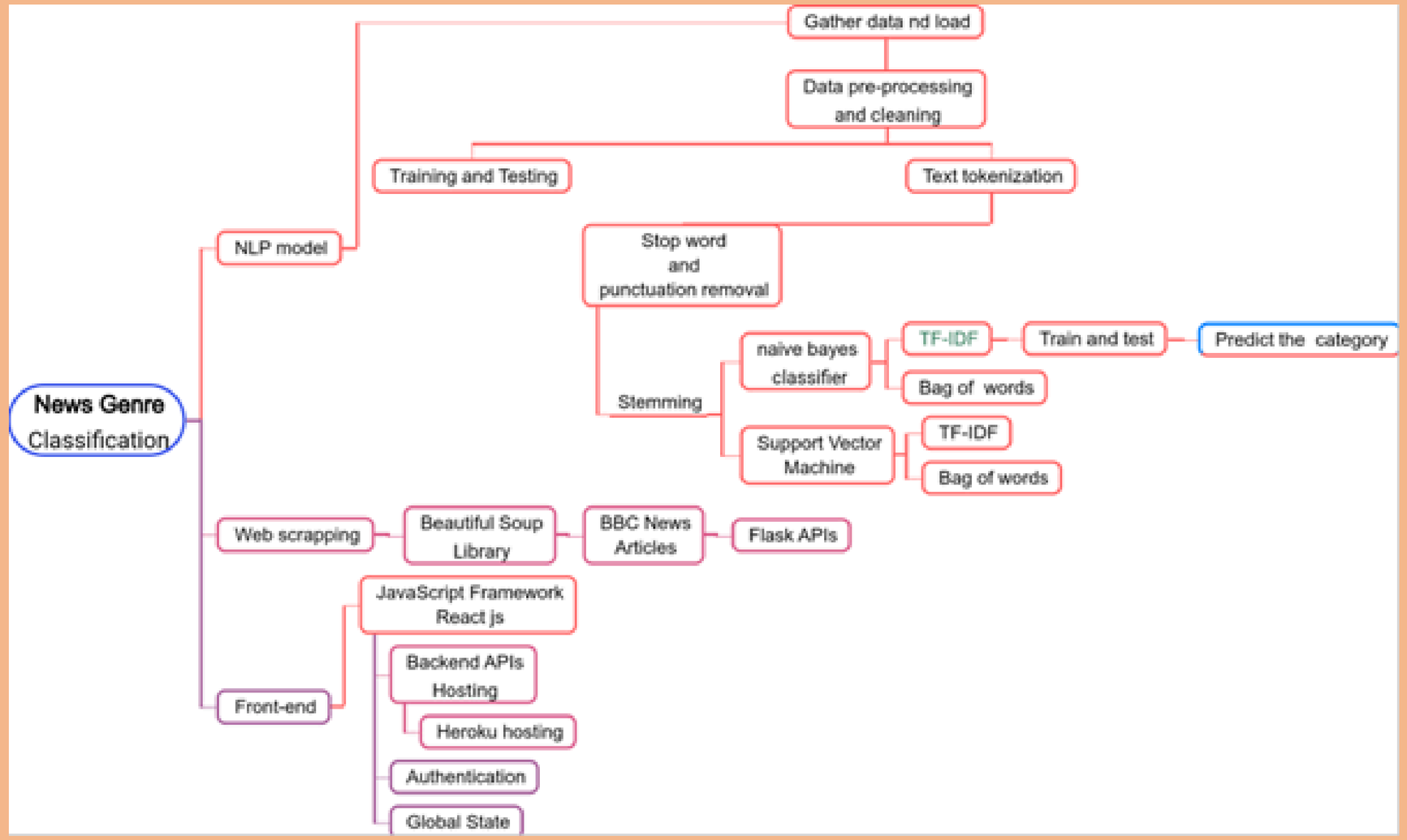
- tf-idf (Term frequency - inverse document)
- BOW (Bag of words)

Libraries

- Pandas
- Seaborn
- Wordcloud
- Nltk
- Sklearn

Flow Chart


News Genre Classification
Landing Page
Sign up
Login
Home page
Latest News
Buttons
Business
Politics
Sports
Entertainment
Top news
Profile
Account settings
Sign-up
Login
Logout
Mini-classifier
Navbar
Top News
Classifier
Category
Top News
Technology


News Genre Classification
NLP model
Gather data nd load
Data pre-processing and cleaning
Training and Testing
Text tokenization
Stop word and punctuation removal
Stemming
naive bayes classifier
TF-IDF
Train and test
Predict the category
Bag of words
Support Vector Machine
TF-IDF
Bag of words
Web scrapping
Beautiful Soup Library
BBC News Articles
Flask APIs
Front-end
JavaScript Framework React js
Backend APIs Hosting
Heroku hosting
Authentication
Global State

Timeline

Milestone 1

- Selecting dataset
- Front-end Designing

Milestone 2

- Training model

Milestone 3

- Fetching news using web scarping

Milestone 4

- Integrating model with web application

References

Books

- Natural Language Processing in Artificial Intelligence by Brojo Kishore and Raghvendra Kumar
- Natural Language Processing with python and spacy by YuliVasilev

Links

- A Comprehensive Guide to Understand and Implement Text Classification in Python (analyticsvidhya.com)
- pandas - Python Data Analysis Library (pydata.org)
- Natural Language Processing With spaCy in Python – Real Python
- spaCy · Industrial-strength Natural Language Processing in Python
- Natural Language Toolkit — NLTK 3.6.2 documentation
- scikit-learn: machine learning in Python — scikit-learn 0.24.2 documentation

THANK YOU!!