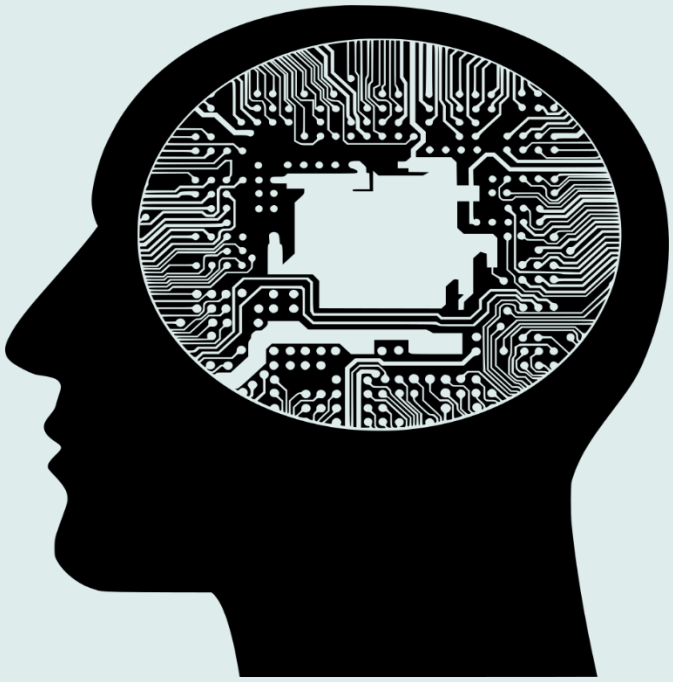
Date:26/09/2023
Course: AI and ML

PHASE 1 SUBMISSION



Ambalavanan N

RegNo: 820421104006

Project Topic: Fake News Detection using NLP (Natural Language Processing)

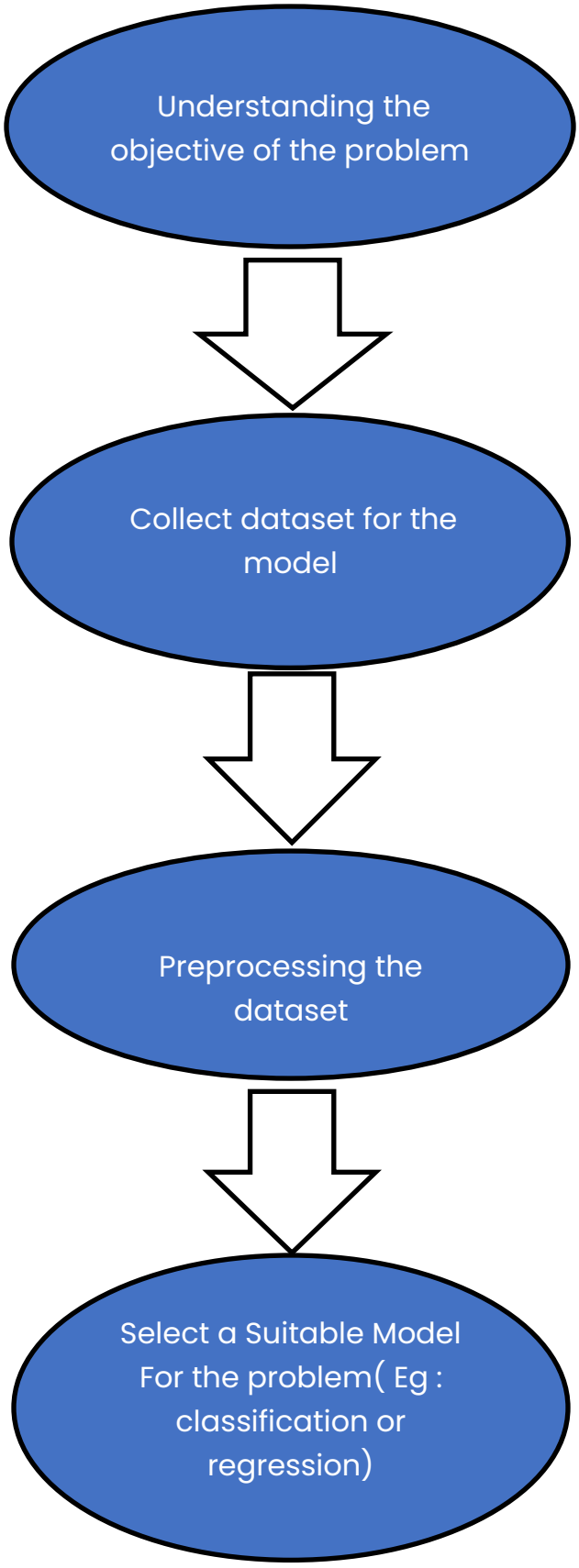
Problem Statement :

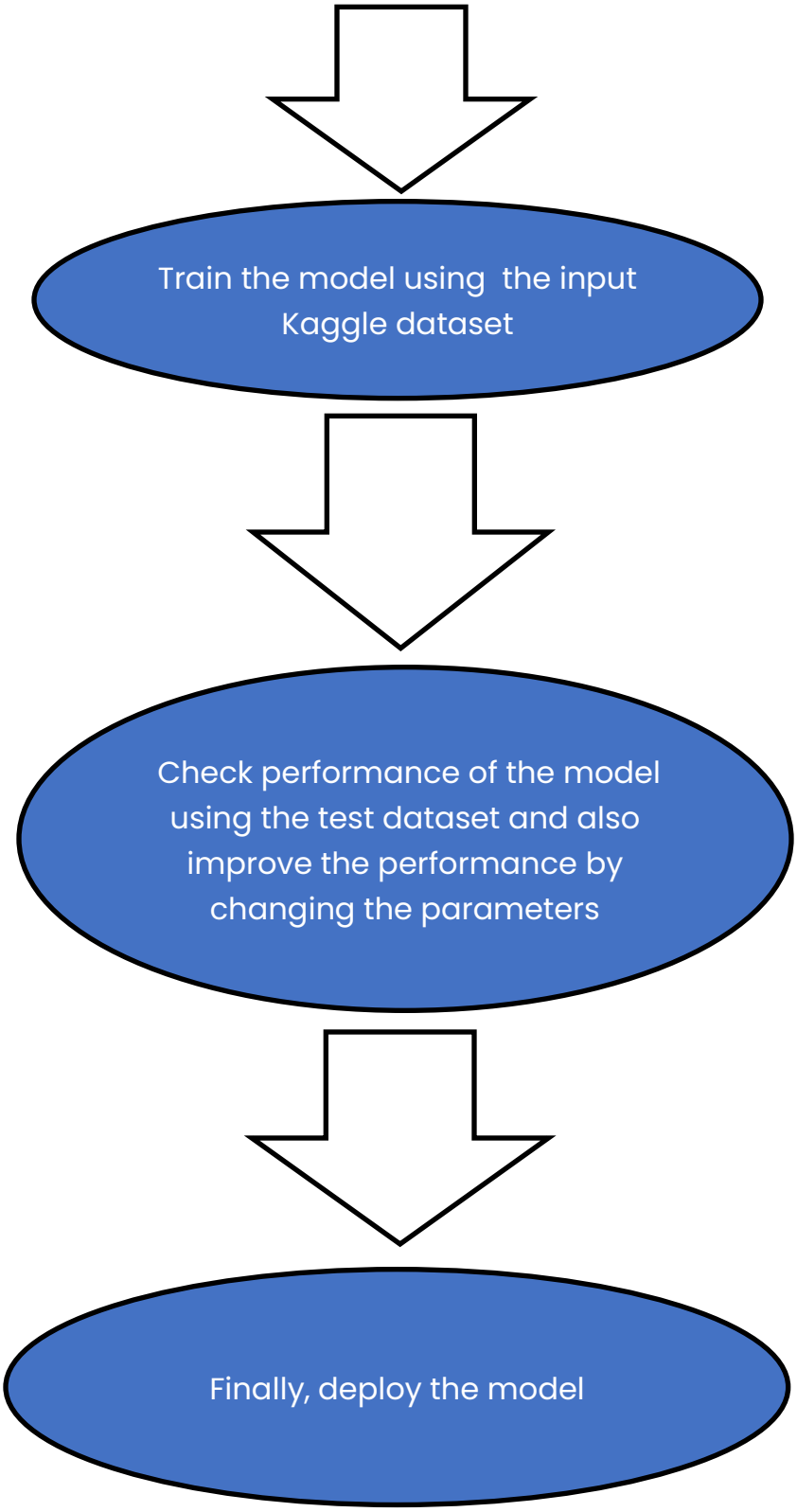
The problem is to develop a fake news detection model using a Kaggle dataset. The goal is to distinguish between genuine and fake news articles based on their titles and text. This project involves using natural language processing (NLP) techniques to preprocess the text data, building a machine learning model for classification, and evaluating the model's performance.

Understandings from the problem statement:

- The problem assigned to me is to create a machine learning model using the NLP(Natural Language Processing) techniques to detect whether a news is genuine or a fake news.
- Here, I advised to use a Kaggle dataset which is used to train the machine learning model.
- Next, I have make a step-by-step process to accomplish the solution for the given problem statement.

Step-by-Step Progress:




Train the model using the input Kaggle dataset
Check performance of the model using the test dataset and also improve the performance by changing the parameters
Finally, deploy the model

Understanding the objective of the problem:

Given problem statement is to create a machine learning model using the NLP(Natural Language Processing) techniques to detect whether a news is genuine or a fake news.

So, here we are going to use a classification model because we just want to classify whether the news is fake or not using their labels. It can be done only using the classification technique.

The classification model will be our choice whether it is a naïve bayes classification model **or** logistic regression (since it is a logistic regression we can use it because it is a best model for binary classification) **or** Neural Networks **or** Random Forest **or** SVM(Support Vector Machine) **or** Word Embedding techniques.

Data Collection:

In this phase, we just collect the data from the Kaggle site.

Data Preparation:

Data which we collected may or may not be accurate or some columns may have missing values.

So, for that we just have to preprocess the data using sklearn or using the pandas commands to remove the unknown and incorrect values.

And also remove the unnecessary column numbers .Change the text format data into numerical format if the model does not compatible with the text format.

Select a Suitable Model:

Most important thing is the model selection. We have to select the correct model for our problem based on our needs.

So , for that we have to test a pool of models then select the best from that, because a model may be a best choice by its overall performance but it may not be best for our problem.

Train the model:

Train the model using the dataset by separating a segment of data for the testing because a model may give best results when we use the same training for the testing.

So, to avoid this we just need to separate two segments training set and testing set. It can be done by using the sklearn commands.

Performance Evaluation:

After we tested our model, we have to just check the performance of our model and try to improve the performance by changing the parameters such test size, checkpoints, echoes, and sometimes we have to make some changes in the preprocessing so that only we get optimized results.

Final Deployment:

After completing all phases mentioned above we can deploy our Fake news detection model