Hate Speech Detection in Arabic Text
Report(1)

Name : Mohamed Ahmed Mohamed Eissa
ID : 20100303

Abstract

This research explores various models for detecting hate speech in Arabic text. The study assesses classical machine learning algorithms, advanced deep learning techniques, and Transformer models. The objective is to develop a robust system for hate speech detection, addressing the unique challenges posed by the Arabic language's complex morphology, rich vocabulary, and various dialects.

1. Introduction

1.1 Problem Statement

The rise of social media has significantly increased the prevalence of harmful content online, making hate speech detection essential for maintaining respectful digital spaces. Detecting hate speech in Arabic is particularly challenging due to the language's complex morphology, rich vocabulary, and various dialects.

1.2 Objective

To develop a robust system for detecting hate speech in Arabic text, leveraging classical machine learning, advanced deep learning, and Transformer models.

1.3 Foundation

This research is based on the GitHub repository Hate-Speech-Detection_OSACT4-Workshop.

2. Progress report :

Completed Tasks:
2.1 Dataset:

Data preprocessing :
Firsy we do some preprocessing technique to make input text clean
Preprocesssing function we use
(Remove diacritics ,Normalize arabic ,Remove punctuations,
Remove repeating char , remove english word and numbers
clean_space)

2.2 Data Augmentation :

Data augmentation is a technique used in deep learning to increase the diversity and size of a training dataset without actually collecting new data.
There are many ways to do this such as:

Synonym Replacement: Replacing words with their synonyms.

Random Insertion: Inserting random words at random positions.

Random oversampling: increasing the number of the small class by randomly duplicating existing instances.

we use Random oversampling

3.0 Data preparation

The model used in this research was classifying whether the text was Hate speech or Not_hate Speech, so I made modifications to the model I used to make it classify the text as either Hate speech or not_hate speech , and also as offensive or not offensive.

4.0 Model Classification

4.1 Classical Machine Learning Algorithms with tf-idf

We do 3 classical Machine Learning Algorithms :

4.1.1 SVM accuracy = 95 Recall = 50

finds the hyperplane that best separates data into different classes.

4.1.2 Random Forest accuracy = 96 Recall = 43

An ensemble method that uses multiple decision trees to make predictions.

4.1.3 Logistic Regression accuracy = 95 Recall = 52

A linear model used for binary classification.

4.2 Deep Learning Models with AraVec Word Embeddings

We do 3 Deep learning model :

4.2.1 Long short Term Memory accuracy = 96 Recall = 55

A type of RNN that can capture long-term dependencies in sequential data.

4.2.2 Gated Recurrent Unit accuracy = 95 Recall = 55

A variant of LSTM with a simpler architecture and fewer parameters.

4.2.3 convolution neural network accuracy = 97 Recall = 57

Primarily used for image data, but can be applied to text data by treating it as a sequence of characters or words.

4.3 Transformers Models with word Embadding

We do 2 Transformers model :

4.3.1 Arabert accuracy = 92 Recall = 80

Bidirectional Encoder Representations from Transformers, and is

designed to capture the context of words in a sentence by considering the words that come before and after it.

4.3.2 marbert accuracy = 93 Recall = 85

In Progress Tasks :

5.0 Data Preprocessing

5.1 we add a function to delete emoji or replace it with a word

6.0 Models

6.1 Classical Machine Learning : will add more classical machine learning model

6.1.1 Decision Trees

A decision tree is a flowchart-like structure where each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes).

6.1.2 Gradient Boosting

Gradient Boosting is an ensemble technique that builds models sequentially, each new model correcting errors made by the previous ones. It combines the predictions of several base estimators to improve robustness.

6.2 deep Learning Models : Will Add BLSTM Deep Learning model

6.2.1 BLSTM

BLSTM networks are an extension of LSTM networks that can process data in both forward and backward directions. This allows the network to have both past and future context, which is particularly useful for tasks where context from both directions is important.

6.3 Transformers Models

6.3.1 ARAT5

T5 is a transformer-based model designed to handle a wide range of NLP tasks by converting them into a text-to-text format

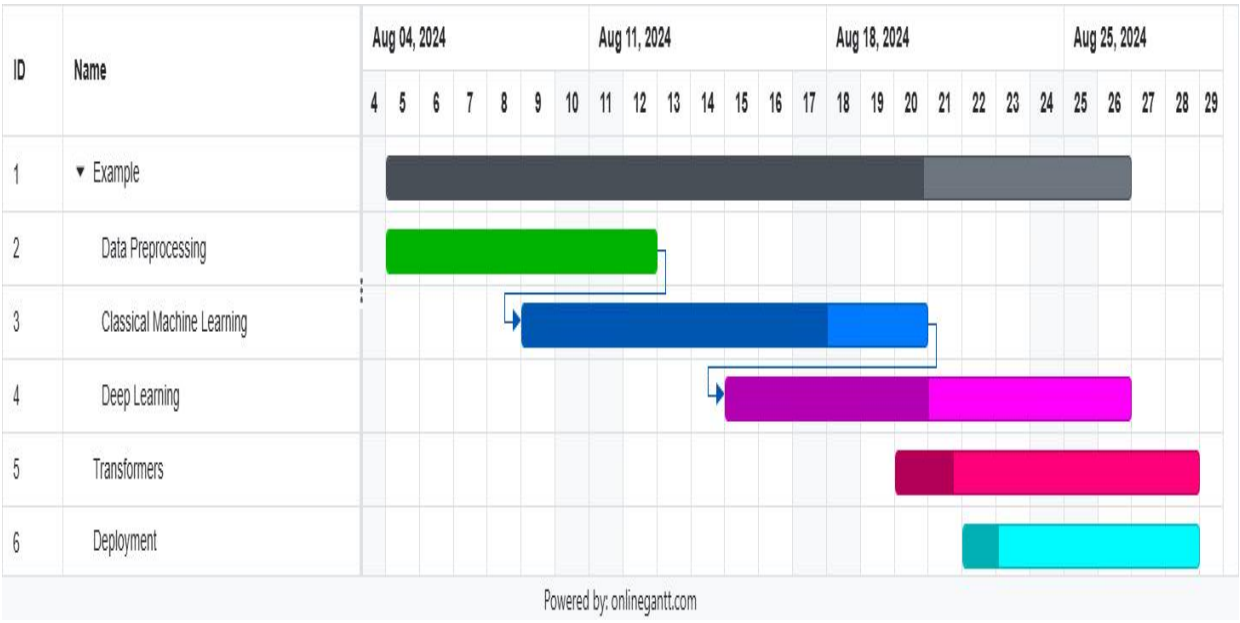
6.3.2 ARAGPT2

GPT-2 is a transformer-based model trained to predict the next word in a sequence

6.3.3 EmoRoBerta

RoBERTa is an optimized version of BERT (Bidirectional Encoder Representations from Transformers) with improvements in pre-training strategies, making it more effective for understanding language context.

7.0 Deployment : Deploying models like ARABERT, MARBERT, and ARAT5


ID
Name
Aug 04, 2024
Aug 11, 2024
Aug 18, 2024
Aug 25, 2024
4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
1 Example
2 Data Preprocessing
3 Classical Machine Learning
4 Deep Learning
5 Transformers
6 Deployment
Powered by: onlinegantt.com