Emotion Analysis 22/23

# Emotion Classification

AKSHAT GUPTA
TEJASWI CHOPPA

DATE: 28TH JANUARY 2023

# Introduction

- Task: Classify emotions in text data based-on Ekman's emotions.
- Motivation : Understand emotions of people to acknowledge their behavior
- Datasets: ISEAR, Emotion Dataset for NLP
- Models: NRCLex and BERT
- Evaluation: Precision, Recall, and F1
- Analysis: Limitations of Lexicon and BERT model.

# Data

## ISEAR (International Survey on Emotions Antecedents and Reactions)

- Data Source : https://www.unige.ch/cisa/research/materials-and-online-research/research-material/
- Type of Data : Student respondent data across the globe to various situations
- Labels present : 6 (joy, fear, anger, sadness, shame, and guilt)
- License: CC-by-sa

## Emotion Dataset for NLP

- Data Source : https://www.kaggle.com/datasets/praveengovi/emotions-dataset-for-nlp
- Type of Data : Tweets dataset
- Labels present : 8 (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust)
- License: CC-by-sa

# Methods

## Lexicon-based approach

- Performed tokenization, cleaning of HTML tags, and lemmatization on both the datasets
- Used the NRCLexicon package by Saif M. Mohammad
- Evaluation using Precision, Recall and F1 score

## BERT-based Sequence Classification

- Performed tokenization using BERT tokenizer
- Applied BERT Sequence Classification using transformers library
- Evaluation using Precision, Recall and F1 score

# Experimental Setup

## ISEAR Dataset

- Total number of instances : 6921
- 80:20 Train - Test split
- Dropped shame and guilt

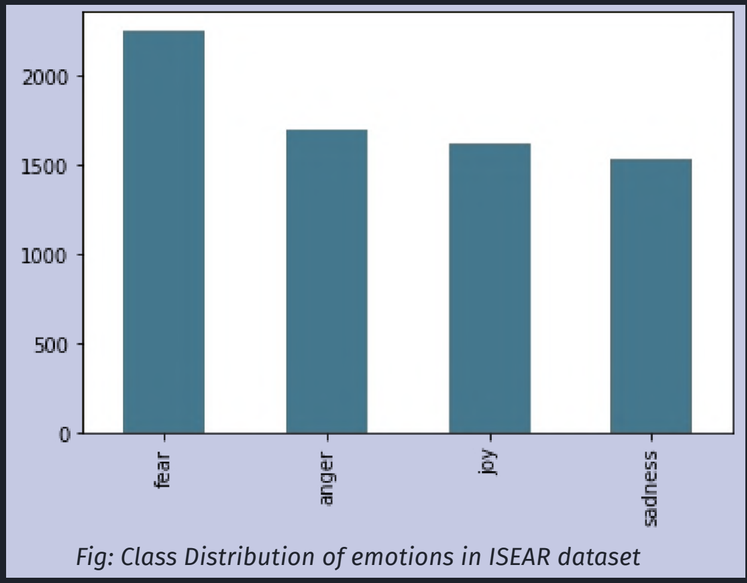


*Fig: Class Distribution of emotions in ISEAR dataset*

## Emotion Dataset for NLP

- Total number of instances : 20k
- Train :  16k , Dev : 2k  , Test : 2k
- Dropped 2 emotions (anticipation and trust)

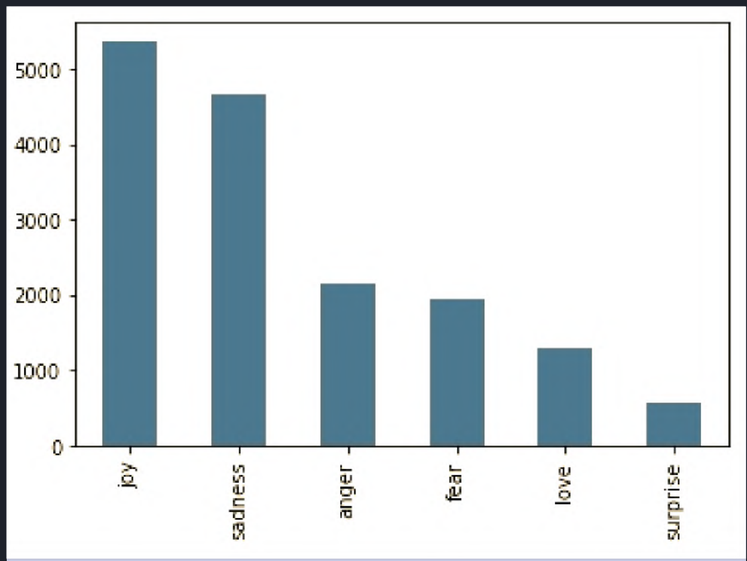


*Fig: Class Distribution of emotions in Emotion dataset for NLP*

# Results

## Lexicon-based approach

- For ISEAR Dataset:

  Precision : 0.37 , Recall : 0.29 , F1 Score :0.33

- For Emotion Dataset for NLP:

  Precision : 0.36 , Recall : 0.27 , F1 Score: 0.31

## BERT-based Sequence Classification

- For ISEAR Dataset:

  Precision : 0.95 , Recall : 0.93 , F1 Score : 0.94

- For Emotion Dataset for NLP:

  Precision : 0.92 , Recall : 0.93 , F1 Score: 0.91

# Analysis

- For Uniformity, subset of Ekman's emotions were used
- With Lexicons, we get POS tags which gives semantic information
- Semantic information works as a feature for classification model
- In Lexicon approach, most of the instances were misclassified
- Negations and OOV(Out of Vocabulary) couldn't be handled in Lexicon approach
- For text classification, BERT produces better results than Lexicon approach
- Bert takes sequential information which is more representative than individual words

# Thank you