Open IIT Data Analytics

TEAM DA11

Index

1

Introduction

Task

Given a pair of sentences, our task is to assess whether they are similar or not based on some form of Gender Bias.

Natural Language Processing (NLP)

Textual similarity is one of the essential techniques of NLP which helps us to measure the extent of similarity between a given pair of text fragments.

Gender Bias

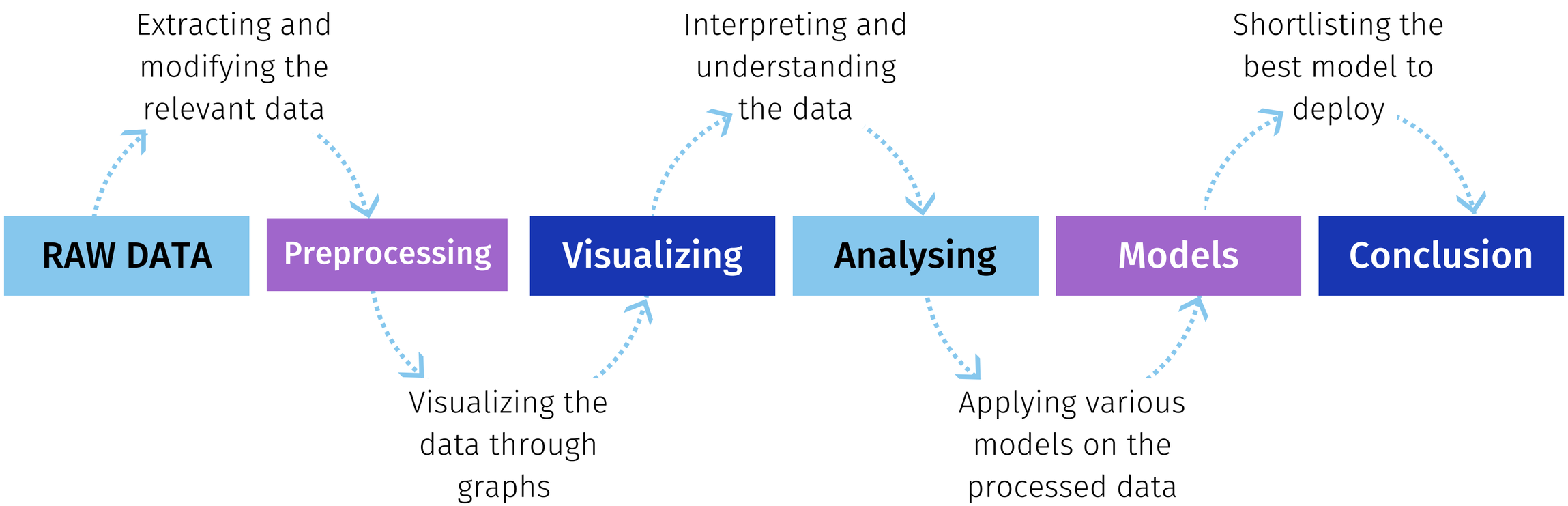
We often see some gender bias in our day-to-day sentences, which could be based on:

- Firstness
- Stereotype
- Subordination

2

Workflow

Workflow


Extracting and modifying the relevant data
Interpreting and understanding the data
Shortlisting the best model to deploy
RAW DATA
Preprocessing
Visualizing
Analysing
Models
Conclusion
Visualizing the data through graphs
Applying various models on the processed data

3

Data Pre-Processing

T

Tokenization

Dividing the text materials into tiny bits. Tokens help in comprehending and developing better models.

P

Punctuation Removal

Punctuation is also turned into a token using tokenization. These tokens must be taken away.

L

Lemmatization

Converts all tokens into their respective base words to avoid complications.

S

Stopwords Removal

Removal of common words like a, an, and, but etc., to make the data precise.

V

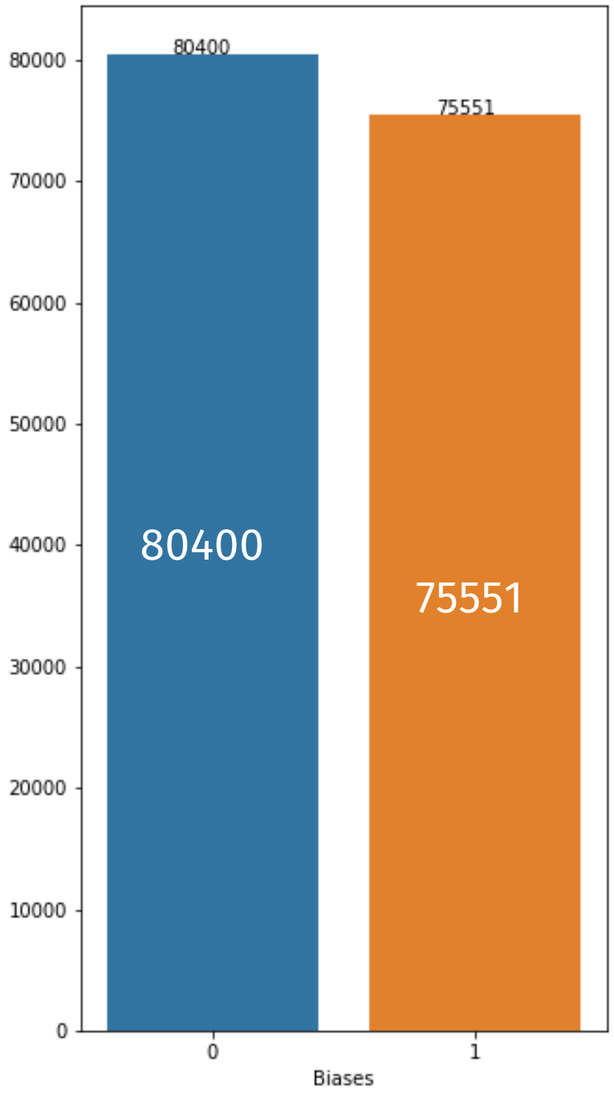
Vectorization

Using TensorFlow's built-in tokenizer module to create an array of words to a vector.

4

Exploratory Analysis

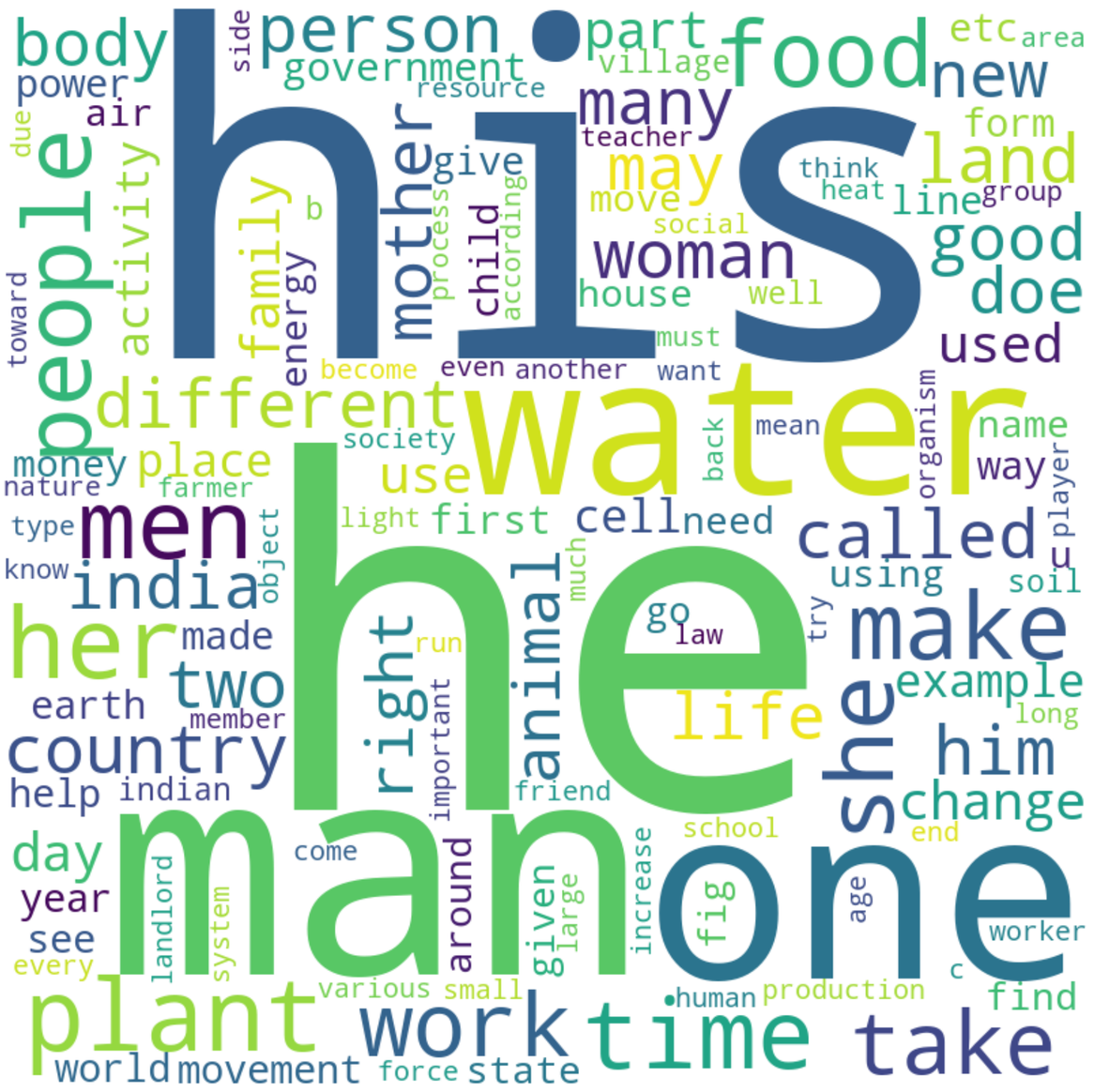
Frequency of labels



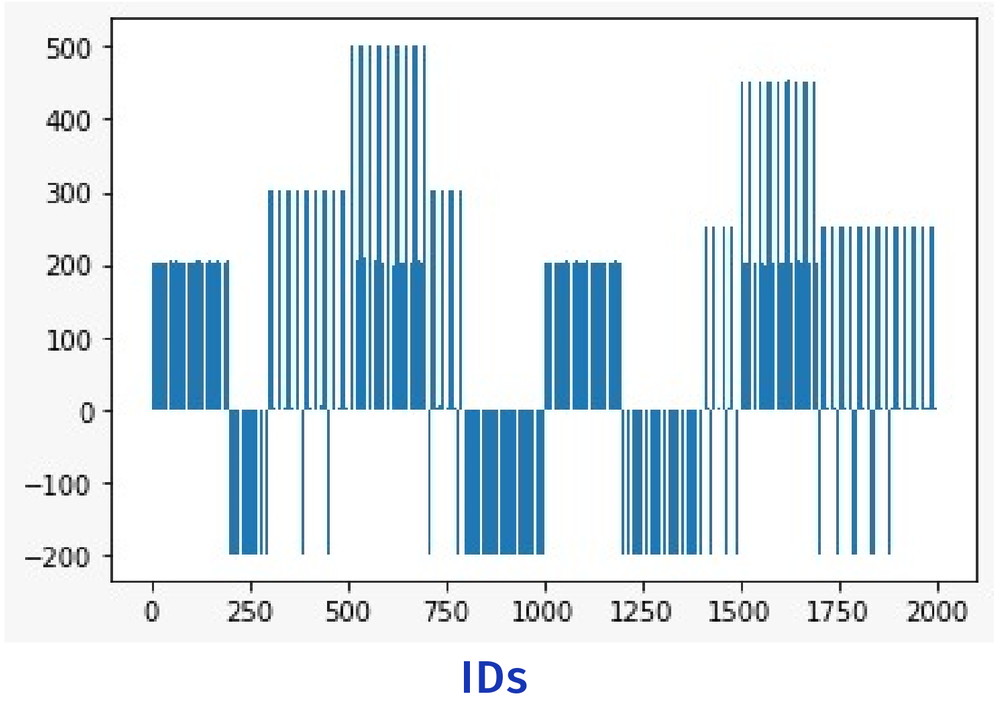
- A plot of number of times 1 and 0 appears in the given training data.
- From here we could infer that the training set had an almost equal number of similar and non-similar pairs of text.

Wordcloud

- Frequently appearing words are depicted in a larger size.
- High frequency of gender depicting words signify presence of gender bias in the set.



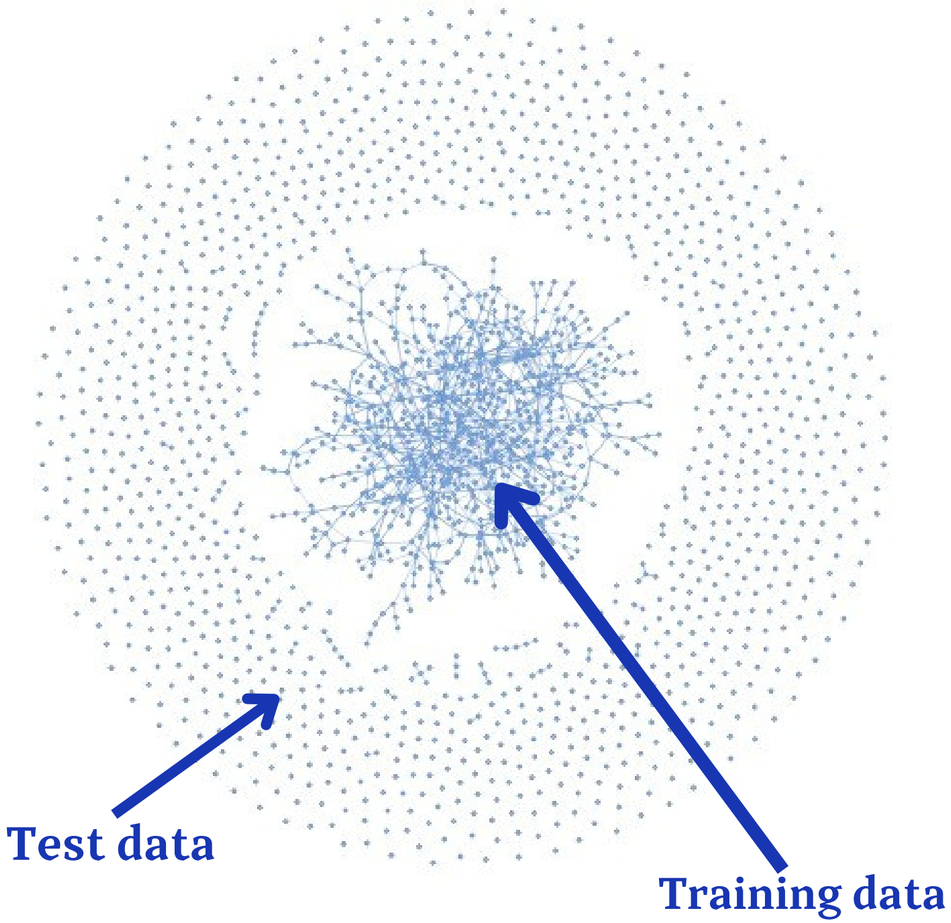
Frequency of IDs



- Plot of how often an ID occurs in the train dataset.
- Negative value of frequency depicts absence of that ID in the train dataset.

Connections

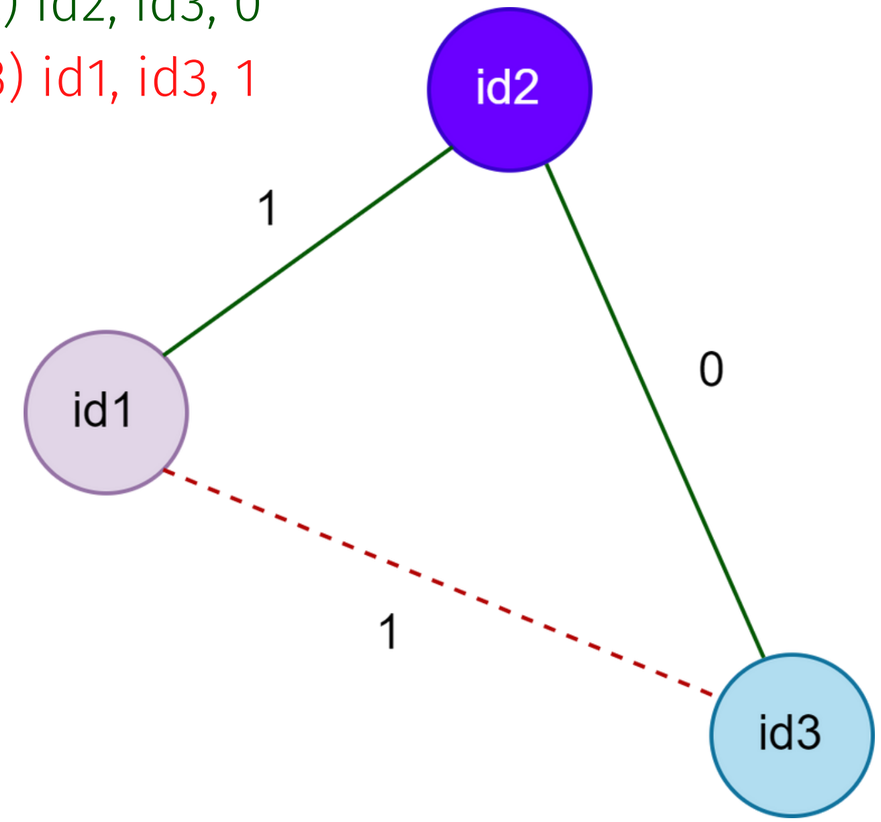
- A graph is constructed with each statement ID as a node and each training example as an edge
- Since no ID from the test data appears in the training instances, the test IDs are fully disconnected.
- To create a proper training and validation dataset, a disjoint training and validation split is made.



Statement Analysis

Training Examples

1) id1, id2, 1
2) id2, id3, 0
3) id1, id3, 1



- Examples 1 and 2 are sufficient to determine the similarity between ID1 and ID3, as one is similar to ID2 and the other is not.
- As the training data form a single connected component we can find similarity between every pair of statements using Breadth First Search on the graph.

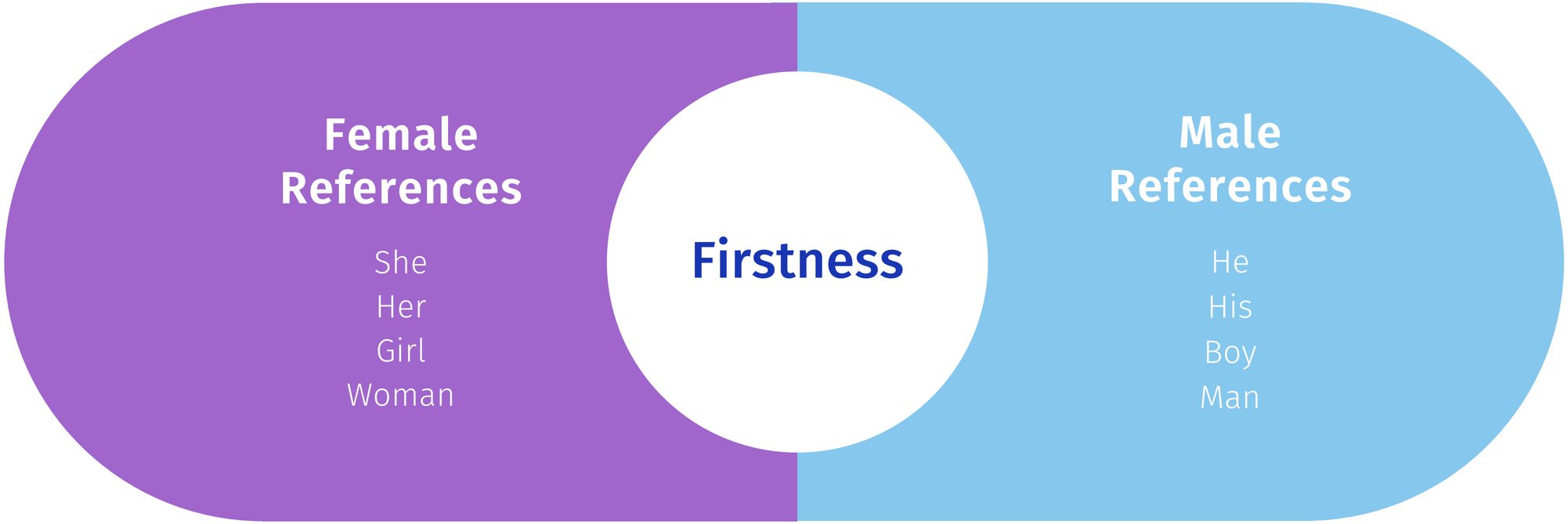
5

Approach and Models

Firstness Model

The model predicts if a statement is biased or not based on the occurrence of gender depicting words. Presence of a male reference before a female reference has been marked as biased. Example:

- 48 Both young boys and girls take part in religious and social functions .



Subordination Model

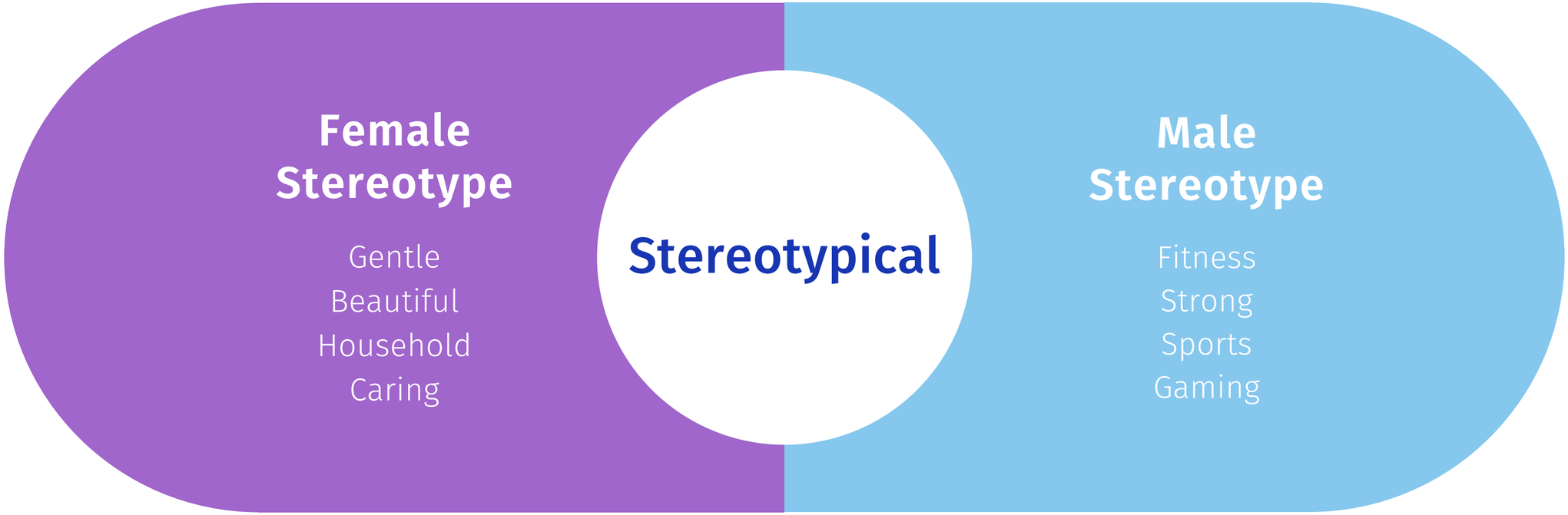
The model predicts if a statement is biased or not based on the occurrence of a comparison word, provided the sentence contains a reference of both genders.The presence of a negation word along with a comparison word would make the sentence unbiased. Example:

- 477 In boys, the muscles of the body grow more prominently than in the girls. Thus, changes occurring in adolescent boys and girls are different.

Stereotype Model

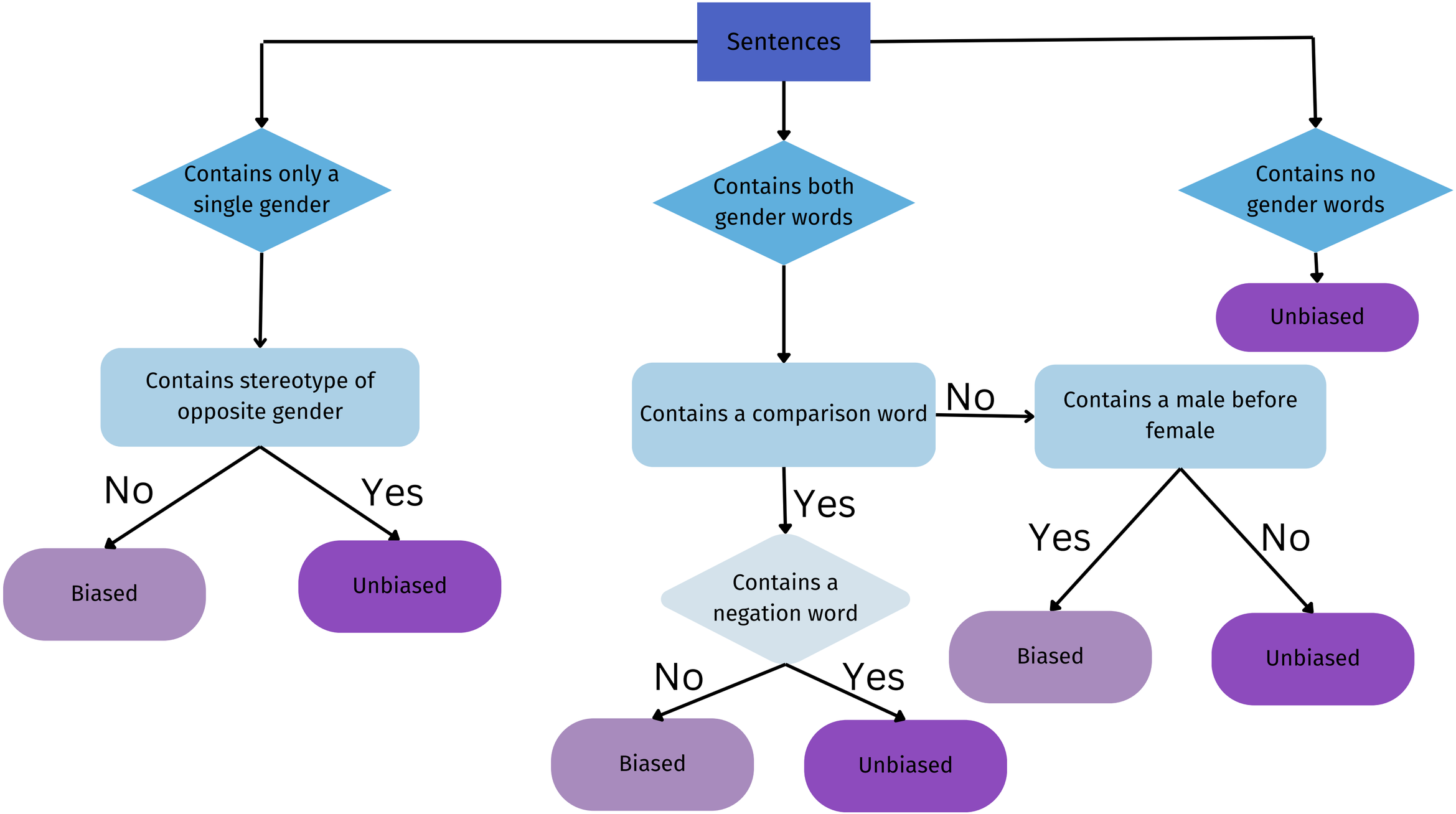
The model predicts if a statement is biased or not based on the presence of a stereotypical reference. A sentence with the same stereotypical reference as that of the gender present in it, is marked as biased. Example:

- 366 Kitchen is the working place for the women.



Combined Model

- Gender bias has been evaluated on individual parameters of Firstness, Subordination and Stereotypical in their respective models
- These 3 parameters are put together to give us the Combined model.
- The flow chart of the Combined model is mentioned in the next slide


Sentences
Contains only a single gender
Contains both gender words
Contains no gender words
Unbiased
Contains stereotype of opposite gender
No
Yes
Biased
Unbiased
Contains a comparison word
No
Contains a male before female
Yes
Contains a negation word
Yes
Yes
No
Biased
Unbiased
No
Yes
Biased
Unbiased

Accuracy

Model	Accuracy
ANN	64%
RNN	79%
Cosine Similarity	67%
Firstness	84.1%
Subordination	84.7%
Stereotype	84.5%
Combined Model	88%

6

Final Approach

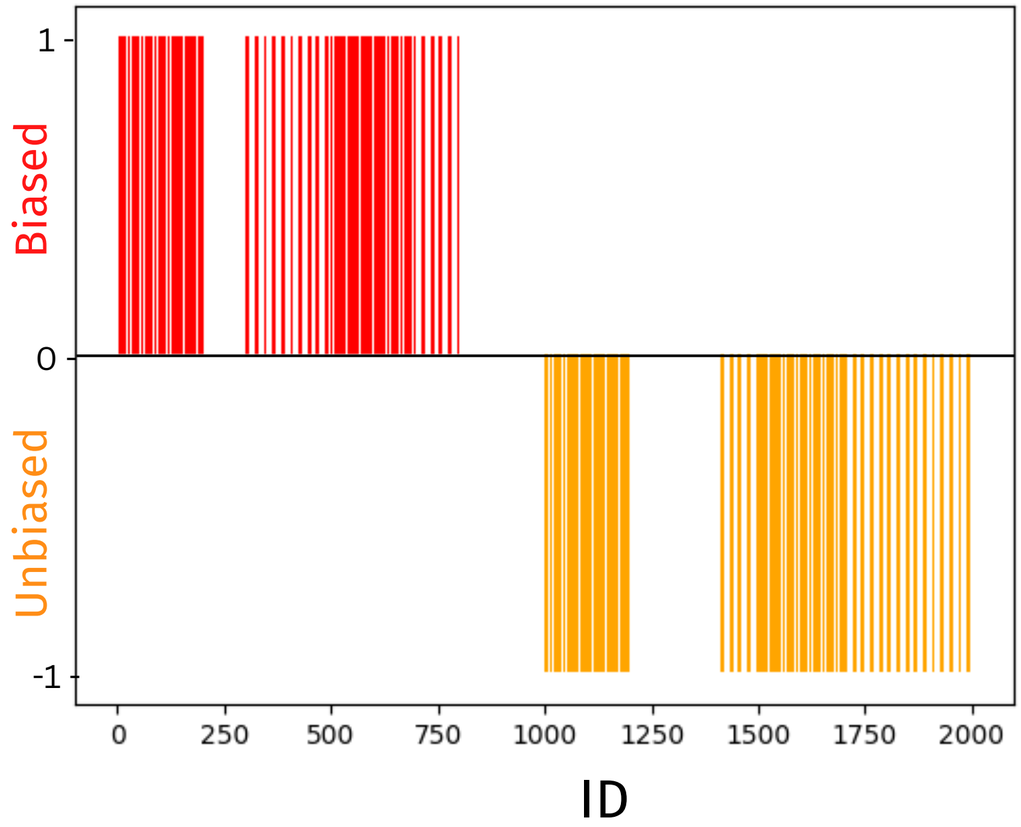
Final Approach

An interesting observation was made in graphical analysis of training datset.

We chose a statement manually and assigned it as biased and ran a BFS on the connections graph

Statements with ID 1-1000 were biased, and statements with ID 1001 to 2000 were unbiased.

This model gave us an accuracy of 100%.



THANK YOU!