

## Data Collection and Preprocessing Phase

Date	03 November 2024
Team ID	739730
Project Title	Figurative Intelligence: Machine Learning for Simile and Metaphor Detection
Maximum Marks	6 Marks

### Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description
Data Overview	"Figurative Intelligence: Machine Learning for Simile and Metaphor Detection," aims to bridge this gap by leveraging advanced machine learning algorithms to accurately identify and classify similes and metaphors.
Univariate Analysis	Univariate analysis is understanding the data with a single feature.
Bivariate Analysis	Bivariate analysis explores the relationship between two variables simultaneously. It investigates how changes in one variable correlate or interact with changes in another variable
<b>Data Preprocessing Code Screenshots</b>	
Loading Data	<pre>df=pd.read_csv(r'C:\Users\Thanuja\Desktop\Data.csv')</pre>

## Handling Missing Data

```
df.shape
```

```
[4]
```

```
... (391, 3)
```

```
df.info()
```

```
[5]
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 391 entries, 0 to 390
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0    Id          391 non-null    int64
1    Sentence    391 non-null    object
2    Target      391 non-null    object
dtypes: int64(1), object(2)
memory usage: 9.3+ KB
```

```
df.isnull().sum()
```

```
[6]
```

```
... Id          0
Sentence      0
Target        0
dtype: int64
```

## Save Processed Data

```
import pickle
with open('vector_model.pkl', 'wb') as file:
    pickle.dump(vectorizer, file)
```

```
[1]
```

```
with open('label_encoder.pkl', 'wb') as file:
    pickle.dump(label_encoder, file)
```

```
[2]
```

```
model.save('trained_model.h5')
```

```
[3]
```