



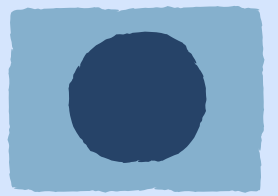
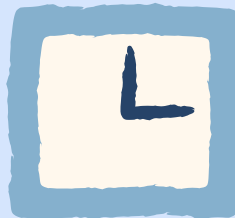
Sarcasm Detector

NLP Project

Cracow University of Technology

Department of Computer Sciences





Contents

- Introduction
- Theoretical Part
- Practical Part
- Summary



Introduction

AIM

Develop and evaluate various machine learning and deep-learning models for detecting sarcasm in text

SCOPE

- Data Collection
- Data Preprocessing (EDA)
- Model Implementation
- Supervised learning algorithms
- Neural network models
- Transformer-based models
- Model Evaluation


METHODOLOGY

- Using Python in Jupyter Notebook and Google Colab
- Kaggle

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from os import path
sns.set()
%matplotlib inline
import time
import calendar
from wordcloud import WordCloud
from collections import Counter
import re
import nltk
import string
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
import datetime as dt
from sklearn.pipeline import Pipeline
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import make_pipeline
from sklearn.model_selection import cross_val_score
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score, confusion_matrix, recall_score,
```

Theoretical Part

In our project, we employed three different types of machine and deep learning models to achieve our objectives.

	Models		
Supervised Learning algorithms	Gradient Boosting	Logistic regression	Too long to train and not the Best
Neural network models	LSTM (Long Short-Term Memory)	GRU (Gated Recurrent Unit)	Too long to train and not the Best
Transformer-based models	RoBERTa (Robustly optimized BERT approach)		The best one

Logistic Regression	Simplicity and Interpretability	Low Computational Cost	Baseline Performance
Gradient Boosting	High Predictive Accuracy	Feature Importance	Robustness to Overfitting
LSTM (Long Short-Term Memory)	Handling Long-Term Dependencies	Sequential Data Processing	Flexibility with Various Sequence Lengths
GRU (Gated Recurrent Unit)	Efficiency	Effective Sequence Modeling	Reduced Complexity
RoBERTa (Robustly optimized BERT approach)	Pre-trained on Extensive Data	State-of-the-Art Performance	Contextual Understanding

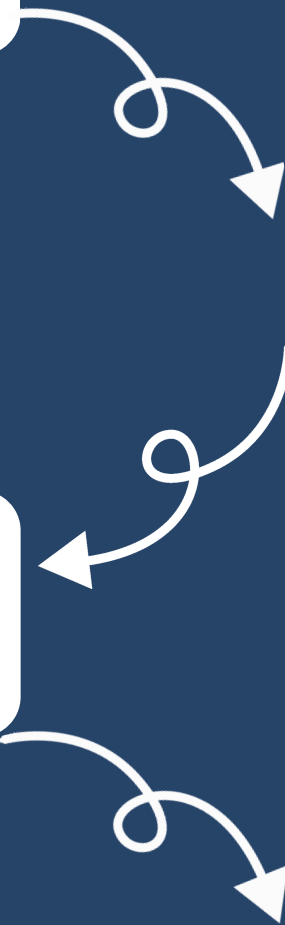
Practical Part

Data collection

Data Preprocessing

Model Implementation

Model Evaluation



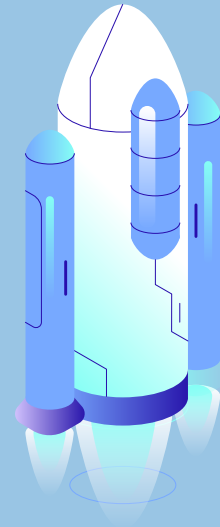
CONCLUSION

In conclusion, this project enabled us to apply the knowledge we had acquired throughout the semester by developing a project to detect sarcasm in text.

Through the exploration and implementation of our models, it becomes evident that the best model to detect sarcasm in text is the RoBERTa Model.

This project has been a real eye-opener, and we're proud of all the new knowledge and skills we have aquired.





THANK YOU

Thank you for taking the time to learn about our NLP Project. If you have any questions or are interested in getting involved, please don't hesitate to contact us.

Together, let's build the future of digital interaction.

