

Classification of Articles in the BBC Dataset into Sub Categories

Medini Rachapudi

TalentSprint

June 25, 2019

Description

Description

- Aims to classify the articles in the BBC Dataset into sub categories

Why This Project ?

- Good introduction to ML
- Learning as the project progresses

Technology Used

- Python
- Python libraries

Dataset

- BBC Dataset : taken from <http://mlg.ucd.ie/datasets/bbc.html> ¹

¹Research Paper : D. Greene and P. Cunningham. "Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering", Proc. ICML 2006

Dataset

- BBC Dataset : taken from <http://mlg.ucd.ie/datasets/bbc.html> ¹

1 Around 2000 samples

2 Five classes

¹Research Paper : D. Greene and P. Cunningham. "Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering", Proc. ICML 2006

Work Plan

- Data pre-processing

Work Plan

- Data pre-processing
- Feature selection and extraction

Work Plan

- Data pre-processing
- Feature selection and extraction
- Implement classification using one basic algorithm

Work Plan

- Data pre-processing
- Feature selection and extraction
- Implement classification using one basic algorithm
- Extend by implementing other algorithm/s

Progress

- Analysed the dataset

Progress

- Analysed the dataset
- Loaded the data

Progress

- Analysed the dataset
- Loaded the data
- Feature extraction algorithms explored

Progress

- Analysed the dataset
- Loaded the data
- Feature extraction algorithms explored
- Implementation algorithms explored

DISCUSSIONS