Comparison performance of Traditional and machine Learning Approaches using the CoNLL-2003 Dataset.

Project Requirements Document

Project Lead: Anthony Bush Castro   **Supervisor: Manzu Gerald, Ph.D**

Required Software, Libraries and Modules

Programming Language: Python

Instructions:

1. Download and install Anaconda
2. Create a virtual environment using conda "conda create --name nlp"
3. Activate the environment "conda activate nlp"
4. Install the libraries below
5. Use the libraries and modules in jupyter notebook

Libraries:

S/N	Libarary	Installation	Module	Usage
01	Jupyter Notebook	pip install notebook		Run: jupyter notebook
02	Standard Python Libararies (No installation required)		counter	from collections import Counter
			gc	import gc
03	Matplotlib	pip install matplotlib		import matplotlib.pyplot as plt
04	Seaborn	pip install seaborn		import seaborn as sns
05	Numpy	pip install numpy		import numpy as np
06	Datasets	pip install datasets	Load_dataset	from datasets import load_dataset
07	Scipy	Pip install scipy	sparse	from scipy import sparse
08	Scikit-learn	pip install sklearn	LabelEncoder	from sklearn.preprocessing import LabelEncoder
09	CRF	pip install sklearn_crfsuite	CRF	from sklearn_crfsuite import CRF
10	Scikit-learn	pip install sklearn	DictVectorizer	from sklearn.feature_extraction import DictVectorizer
			RandomForestClassifier	from sklearn.ensemble import RandomForestClassifier
			LabelEncoder classification_report confusion_matrix precision_recall_curve precision_recall_curve auc	from sklearn.metrics import classification_report, confusion_matrix, precision_recall_curve, auc