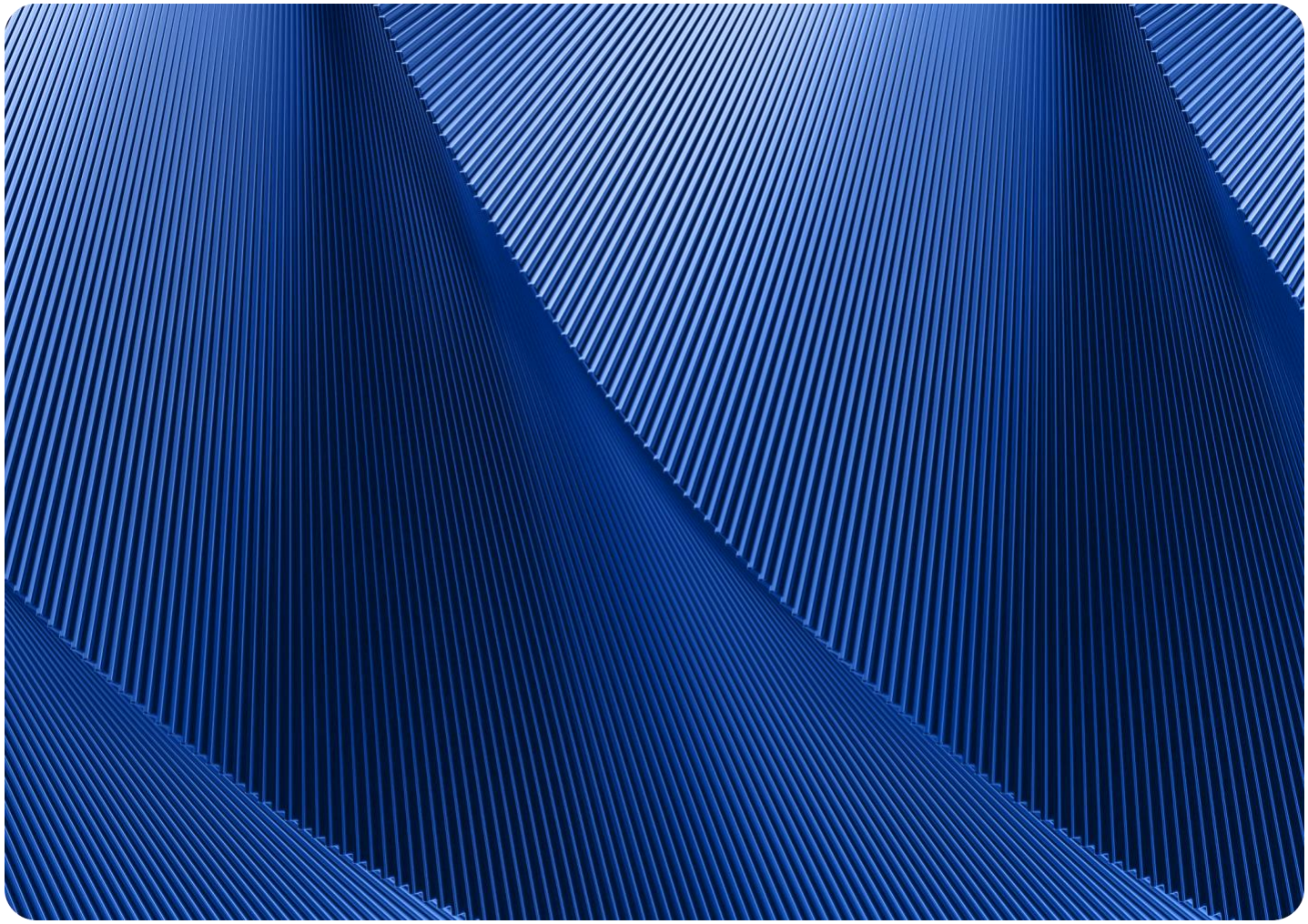
NAAIC

Resource provided by
Miami Dade College

Course Syllabus

Introduction to Natural Language Processing

This introductory course provides hands-on experience with Natural Language Processing concepts, and real-world applications.



NAAIC

Table of Contents

Course Description

Students will learn the fundamental concepts of Natural Language Processing (NLP) and text processing. In addition, focus will be on knowledge and skills necessary to create a language recognition application.

Prerequisites:
Machine Learning Foundations
Introduction to Python

NAAIC

Course Competencies

Competency 1
Students will describe common techniques in Natural Language Processing (NLP) and associated applications by:
a) Exploring AI For NLP, Applications of NLP, NLP data processing, BOW, and Algorithms in NLP b) Processing textual data by sentence segmentation, tokenization, lemmatization, stop word removal, etc. c) Applying data preprocessing techniques like document similarity, Word Vectors, Cosine similarity, etc. d) Distinguishing between NLP models and algorithms
Competency 2
Students will describe the data acquisition process in NLP by:
a) Comparing different types of NLP datasets b) Identifying and examining various data storage methods c) Examining different sources of curated data and interpreting their usage in NLP domain d) Downloading and processing data using the NLTK library e) Applying data visualization techniques specific to NLP
Competency 3
Students will explore NLP Data Preprocessing by:
a) Utilizing proprietary and open-source libraries and data visualization techniques b) Exploring and applying various vectorization techniques c) Exploring and applying the methods of document similarity and vector visualization d) Distinguishing between various distance measurement techniques e) Defining and understanding the various processes associated with the NLP data pipeline

Competency 4
Students will describe, compare, and train different machine learning models by:
a) Describing and applying NLP classifiers to train machine learning models b) Describing neural networks and their working principles c) Understanding various language models d) Defining and summarizing various Neural Language Models, N-gram Models and Sequential Models e) Defining and demonstrating Recurrent Neural Networks and Named Entity Recognition (NER) models through various activities and use cases
Competency 5
Students will explore NLP Model deployment by:
a) Identifying and exploring various machine learning model deployment platforms b) Describing and classifying various types of chatbots by their applications c) Implementing Language Detection, Transliteration, Translation, and Sentiment Analysis for different language scenarios d) Using different tools to create and deploy chatbots, pre-existing chatbot frameworks, and ChatterOn e) Using cosine similarity in neural networks to train chatbots
Competency 6
Students will discuss and describe advanced models in NLP by:
a) Exploring the most recent developments in the NLP space b) Explaining the workings of LSTM, Transformers, and BERT c) Examining several NLP pretrained models d) Comparing the workings, performance, and architectures of DistilBERT, RoBERTa, GPT, GPT -2, BERT, and BART models of NLP e) Summarizing various types of learning fast learning techniques, Zero-shot, One-shot, and Few- shot f) Evaluating various ethical issues in language models

NAAIC

Instructional Resources

No prescribed textbook is required due to the rapidly evolving nature of natural language processing tools, frameworks, and best practices. Instructors will curate up-to-date articles, tutorials, documentation, and other resources available to provide the latest advancements and diverse perspectives. This approach ensures flexibility to tailor content to class needs and incorporate emerging topics and industry practices.

Platforms and Tools:
Kaggle, Google Colab, Jupyter notebook, Orange data mining, Python 3.7+, Anaconda, Intel® OpenVINO 2021

Recommended Reading:

- *Introduction to Natural Language Processing*, Jacob Eisenstein

Grading Schema

Assignment Type	Percentage of Grade
Assignments	20%
Discussions	10%
Projects	60%
Attendance	10%

NAAIC

Course Outline

Week starting on / Module	Module Topic
Week 1	Introduction and Syllabus Overview
Week 2	NLP Algorithms and Data Storage
Week 3	Data Curation; NLP tools
Week 4	Data Visualization
Week 5	NLP Data Preprocessing Part I
Week 6	NLP Data Preprocessing Part II
Week 7	NLP Data Pipeline
Week 8	Introduction to Machine Learning Models
Week 9	NLP Model Deployment
Week 10	Building a Chatbot I & II
Week 11	NLP For Multiple Languages
Week 12	LSTM, Transformers, and ChatGPT
Week 13	Advanced NLP Models
Week 14	Crafting Innovative LLM-Powered Applications