Course Overview:
Introduction to NLP
(Specifically Text Processing)

Muhammad Okky Ibrohim, S.Mat., M.Kom.

Credit

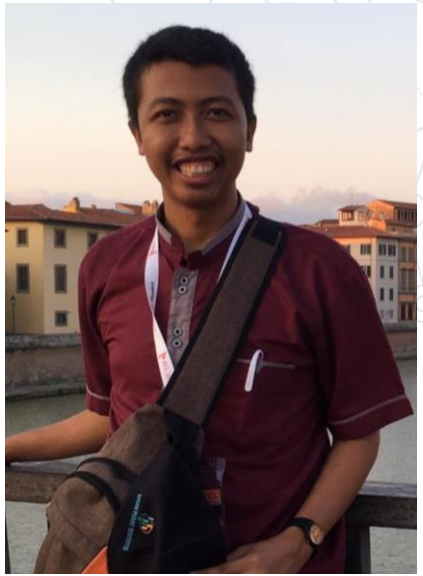
This slide was built using **Blue Connections Cordelia Presentation Template** theme (https://www.slidescarnival.com/cordelia-free-presentation-template/216).

Several images and other media from Google Images that maybe I forgot to add specific credit (if there is belong yours, please contact me if you want me to add specific credit to you or remove it).

Video course

For those who can not attend the synchronous session, you can watch this video for this slide presentation: https://youtu.be/Vmjramv2-2Y

About Me



Full Name and Degree

Muhammad Okky Ibrohim, S.Mat., M.Kom.

Current Position(s)

- PhD student at Università degli Studi di Torino

Previous Relevant Professional Position(s)

- Lecturer and Researcher at Faculty of Computer Science, Universitas Indonesia
- NLP Researcher and Engineer at several start up (part time and project base)

Scholar Link(s)

- Google Scholar: https://scholar.google.com/citations?user=Rh2NplAAAAAJ&hl=en
- ORCID: https://orcid.org/0000-0002-6943-6553

Contact

If you want to ask and/or discuss an NLP project, or want to give advice on this slide/course, feel free to reach me via email (okkyibrohim@gmail.com) or LinkedIn (https://www.linkedin.com/in/muhammad-okky-ibrohim-5683a376/)

Outline

- Prerequisite
- Expected Result
- What you should not expect from this course
- How this course will be delivered
- NLP Research Field (Super Simplified)
- Course Material

Prerequisite

- Basic understanding of math
- Basic understanding of machine learning
- Basic understanding of Python programming
- Basic understanding of Notebook (Jupyter Notebook and/or Google Colab) operation

Expected Result

- Understanding what is NLP and its application
- Understanding fundamental NLP terms
- Understanding fundamental NLP pipeline
- Able to build several NLP task models using best practice way

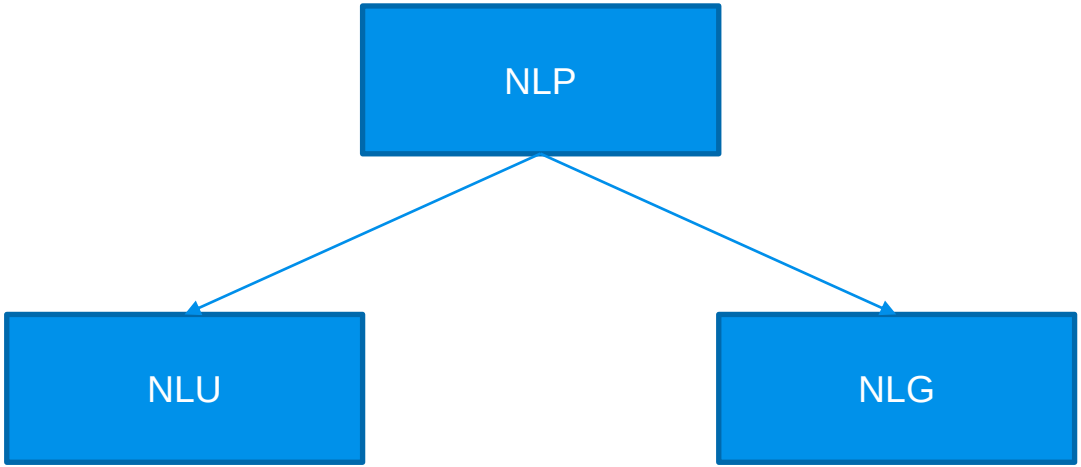
What you should not expect from this course

- Deep discussion about the math behind the NLP algorithms
- Advanced programming example to build NLP architecture from scratch
- Advanced programming example to modify NLP architecture

How this course will be delivered

- Theory
  Discussion of material introduction
- Practice
  Hands-on using Python in Notebook (Jupyter Notebook and/or Google Colab)

NLP Research Field (Super Simplified)



Course Material:
1. Dataset Collection and Building

1.1 Why we need to learn dataset collection and building?

1.2 How to collect dataset

1.3 Data annotation and how to validate it

Course Material:
2. Text Preprocessing

2.1 Why we need to preprocess our dataset?

2.2 Basic text preprocessing in text dataset

2.3 Hands-on: Basic text preprocessing using several NLP libraries (NLTK, Sastrawi, etc.)

Course Material:
3. Feature Extraction

3.1 What is feature extraction?

3.2 Basic text feature extraction in text dataset

3.3 Introduction to feature selection

3.4 Hands-on: Basic text feature extraction and selection (from scratch and using Scikit-Learn library)

Course Material:
4. Text Classification

4.1 Type of text classification

4.2 Text classification pipeline

4.3 Hands-on: Basic document classification using classic machine learning with basic feature extraction (using Scikit-Learn library)

4.4 Hands-on: Basic token classification using classic machine learning with basic feature extraction (using Scikit-Learn library)

4.5 Hands-on: Create text classification library

Course Material:
5. Word Embedding

5.1 Introduction to word embedding

5.2 Training word embedding

5.3 Using pretrained word embedding model from previous work

5.4 Hands-on: Training word embedding (using Gensim and/or other NLP libraries)

5.5 Hands-on: Text classification using pretrained word embedding (using Scikit-Learn and/or other NLP libraries)

5.6 Hands-on: Create text classification library using pretrained word embedding

Course Material:
6. RNN and Its Development

6.1 Introduction to RNN and its development

6.2 Hands-on: Text classification using RNN and/or its development with pretrained word embedding (using TensorFlow and Keras libraries)

6.3 Hands-on: Create text classification library using RNN and/or its development with pretrained word embedding

Course Material:
7. Machine Translation

7.1 Introduction to machine translation

7.2 Hands-on: Training machine translation using NMT and/or other architectures (using TensorFlow and Keras libraries)

7.3 Hands-on: Create machine translation library using NMT and/or other architectures

Course Material:
8. Transformer Model

8.1 Introduction to transformer model

8.2 Hands-on: Training text classification using pretrained transformer model (using Simple Transformers and/or other NLP libraries)

8.3 Hands-on: Create text classification library using pretrained transformer model

Course Material:
9. Case Study NLP

9.1 Aspect Based Sentiment Analysis

9.2 Introduction to Telegram Chatbot

9.3 Bonus: How to conduct research on NLP

Thank You!

Bisa.ai
PT. Bisa Artifisial Indonesia

Kampus Merdeka
INDONESIA JAYA