Group Workshop Syllabus

1. Overview

Title	**Natural Language Processing for Text Classification**		
Prerequisites	**High School Students**	Required course/Knowledge	N/A
		Recommended Materials for preparing for the course	N/A
	College Students	Required course/Knowledge	Python programming Pytorch programming Basic machine learning
		Recommended Materials for preparing for the course	PyTorch tutorial: https://pytorch.org/tutorials/beginner/pytorch_with_examples.html A Course in Machine Learning by Hal Daumé III (http://ciml.info/)

2. Program Introduction and Objectives

Introduction	Natural Language Processing (NLP) is an important topic in Artificial Intelligence with a wide range of applications. Modern NLP is primarily based on statistical methods and machine learning algorithms, where linguistic information is provided by instances of uses of language. This course will cover theoretical fundamentals in NLP but will be mostly guided by a strong practical component: the task of building classifiers using various types of machine learning algorithms (from basic probabilistic models to advanced neural architectures), based on various types of text pre-processing techniques, as well as pre-trained representations, ending with classical ways to evaluate the models.
Course Objectives (Please describe the theoretical and practical objectives of the course)	● Students will gain familiarity with linguistic concepts involved in language understanding and generation, from morphological analysis to pragmatics; ● Students will gain familiarity with, devise,

	implement and apply relevant pre-processing steps for natural language processing components and applications; ● Students will be able to critically compare statistical and deep learning approaches for natural language processing classification tasks; ● Students will be able to build, evaluate, critically analyze and improve models using existing machine learning algorithms and deep learning frameworks (pytorch) for classification tasks from texts.
Software/Tools (if any)	Python and Pytorch as programming languages, Google colab as environment

3. Program Schedule

Week		Lecture
1	**Topic**	Week 1 - Introduction to NLP and pre-processing
2	**Topic**	Week 2 - Text Classification
3	**Topic**	Week 3 - Feed Forward and Convolutional Neural Networks
4	**Topic**	Week 4 - Language Modelling and Recurrent Neural Networks
5	**Topic**	Week 5 - Evaluation and advanced topics in NLP
6	**Final Project Review Week**	
7	**Final Written Reporting and Oral Presentation**	

4. Reading Materials (Incomplete)

Text Preprocessing in Python: Steps, Tools, and Examples (https://medium.com/@datamonsters/text-preprocessing-in-python-steps-tools-and-examples-bf025f872908

Understanding Convolutional Neural Networks for NLP (http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/)

Ian Goodfellow and Yoshua Bengio and Aaron Courville. Deep Learning.Practical Methodology Part. https://www.deeplearningbook.org/contents/guidelines.html

5. Final Project Theme

Develop a classification model to assess Humor in Edited News Headlines. Given the original headline and two edited versions, the participant has to predict which edited version is the funnier of the two. The task is to build an advanced RNN/Transformer model using PyTorch. Final results are to be submitted to the online platform. The evaluation of results is done automatically by the platform.

6. Some Future Research Fields/Direction/Topics

NA