“NLP from Scratch in Azerbaijani” Course Syllabus

Part 1: Introduction to Natural Language Processing

Week 1: Introduction, Text Preprocessing, and Text Representation

▸ Lecture 1: Introduction *[Launch date: 15 Jan]*

- □ Course Structure

▸ Lecture 2: Text Preprocessing *[Launch date: 15 Jan]*

- □ Tokenization: Sentence-level and word-level tokenization
- □ Stemming vs. Lemmatization: Differences and when to use them
- □ Stopword Removal: Why and how to remove irrelevant words
- □ Lowercasing, Punctuation Removal, Special Character Handling: Cleaning text data

▸ Lecture 3: Text Representation *[Launch date: 15 Jan]*

- □ Bag of Words (BoW): Basics of BoW for text representation
- □ TF-IDF: Understanding Term Frequency-Inverse Document Frequency and its applications

▸ Lecture 4: Live Coding: Classifying Azerbaijani Bank Support Service Requests Using TF-IDF *[Launch date: 15 Jan]*

Week 2: Word Embeddings, and Language Models

▸ Lecture 5: Introduction to Word Embeddings *[Launch date: 18 Jan]*

- □ Limitations of BoW and TF-IDF: Sparse vectors and lack of context
- □ Dense Vector Representation: The concept of capturing meaning in embeddings
- □ Word2Vec: Understanding Skip-gram and CBOW architectures
- □ GloVe: Global Vectors for Word Representation

▸ Lecture 6: Live Coding: Training Word2Vec Model for Azerbaijani Dataset Using TensorFlow *[Launch date: 18 Jan]*

▸ Lecture 7: Language Models *[Launch date: 18 Jan]*

- □ Introduction to n-grams: Basics of n-gram language models
- □ Building a Simple n-gram Language Model: Step-by-step guide

Week 3: Text Similarity, Live Coding, Evaluation Metrics and Deep Learning for NLP

▸ Lecture 8: Text Similarity *[Launch date: 20 Jan]*

- □ Cosine Similarity: Measuring similarity with embeddings
- □ Sentence Similarity: Using simple methods for comparison

▸ Lecture 9: Live Coding: Detecting Similar Customer Support Queries *[Launch date: 22 Jan]*

▸ Lecture 10: Evaluation Metrics for NLP *[Launch date: 22 Jan]*

- □ Precision, recall, F1-score, perplexity, BLEU, ROUGE, etc.

▸ Lecture 11: Transitioning from Traditional to Deep Learning for NLP *[Launch date: 25 Jan]*

- □ Why deep learning is essential for modern NLP tasks

Part 2: Advanced Topics in NLP

Week 4: RNNs and Live Coding

▸ Lecture 12: Recurrent Neural Networks (RNNs) *[Launch date: 28 Jan]*

- □ Basics of RNNs for sequence data
- □ Introduction to GRUs and LSTMs to handle long-term dependencies

▸ Lecture 13: Live Coding: Implementing an LSTM Model *[Launch date: 31 Jan]*

- □ Building an LSTM model for text classification

Week 5: Challenges with Sequence-Based Models and Attention Mechanisms

▸ Lecture 14: Challenges with Sequence-Based Models

- □ Vanishing gradients, scalability issues, and long-term dependency challenges

▸ Lecture 15: Introducing Attention Mechanisms

- □ The concept of attention in neural networks
- □ How attention evolved into Transformers
- □ Hands-On: Implementing attention mechanisms in PyTorch

Part 3: Modern NLP Techniques

Week 6: Transformer Architecture, Comparison with Traditional Models and Hugging Face Ecosystem

▸ Lecture 16: Understanding Transformer Architecture

- □ Encoder-decoder framework
- □ Key components: Self-attention, positional encoding, multi-head attention, feedforward layers

▸ Lecture 17: Comparison with Traditional Models

- □ Why Transformers outperform RNNs and CNNs

▸ Lecture 18: Hugging Face Ecosystem Overview

- □ Overview of popular models like BERT, GPT, and T5

- □ Encoder and decoder models

Week 7: Tokenization, and Fine-Tuning for Text Classification

▸ Lecture 19: Understanding Tokenization

- □ Byte Pair Encoding (BPE) and WordPiece tokenization techniques

▸ Lecture 20: Fine-Tuning Transformers for Text Classification

- □ Using Transformers for classification tasks

Week 8: Fine-Tuning for NER and Summarization

▸ Lecture 21: Fine-Tuning Transformers for Named Entity Recognition (NER)

- □ Token classification for NER tasks
- □ Fine-tuning multilingual transformers like XLM-R

▸ Lecture 22: Fine-Tuning Transformers for Summarization

- □ Comparing models like GPT-2, T5, BART, and PEGASUS
- □ Training summarization models with PEGASUS

Week 9: Fine-Tuning for QA and Pretraining RoBERTa

▸ Lecture 23: Fine-Tuning Transformers for Question Answering

- □ Using pre-trained Transformers to build QA systems

▸ Lecture 24: Pretraining a RoBERTa Model from Scratch

- □ Steps to pretrain a RoBERTa model for custom datasets

Part 4: Agentic AI

Week 10-12: Agentic AI

- Lectures 25-32: Agentic AI
  - □ Topics and activities to be determined

1. Lesson Structure

Overview and Learning Objectives

Each lesson begins with a brief overview of the topic, providing context and setting clear learning objectives. This introduction helps students understand the significance of the topic and what they will achieve by the end of the lesson.

Lecture format

Presentation Session: The lecture session delves into the theoretical concepts related to the topic.

Code Session: The code session provides a hands-on approach to learning.

2. Community Engagement

Telegram Group

Community engagement will be facilitated through our Telegram group, where students can join discussions with the authors and other students. The link to join the Telegram group is: https://t.me/+mWr4SrxAeWgwZWMy. This platform encourages a supportive and interactive learning environment where students can ask questions, share insights, and collaborate on projects.

Exclusive Resources

Certain resources, such as homework checking videos, additional tutorials, and GitHub project ideas, will be available only to group members. Joining the group is free and provides access to these valuable resources.

Video Upload Schedule

Videos will be uploaded based on demand and engagement metrics, ensuring that the content is relevant and timely. Each video will be released after a specific number of students join the group or complete a set of tasks. Announcements will be made in the Telegram group when new videos are uploaded. This approach keeps the community engaged and informed about new content.

3. Homework and Projects

Assignments

Homework assignments will be provided to reinforce learning and deepen understanding. These assignments are designed to apply theoretical knowledge to practical scenarios, enhancing the learning experience.

Project Ideas

Both individual and group-based project ideas will be suggested. These projects allow students to apply what they have learned in a more comprehensive and creative manner.

Submission Instructions

Detailed instructions for submitting homework and projects, including deadlines, will be provided. These assignments and projects are integral to the learning process and help students develop practical skills.

Peer Reviews

Students will have the opportunity to review each other's work, providing constructive feedback and learning from different approaches.

4. Q&A Sessions

Live Sessions

Periodic live Q&A sessions will be held to address any questions or doubts. Announcements for upcoming Q&A sessions will be made in the Telegram group. These sessions provide an opportunity for students to interact directly with the instructor, clarify doubts, and gain deeper insights into the topics covered.

Office Hours

Regular office hours will be scheduled for one-on-one discussions with the instructor, providing personalized support and guidance.

Supplementary Materials

Supplementary materials, such as reading lists, research papers, and external links, will be provided to enrich the learning experience. These resources are carefully curated to offer additional perspectives and in-depth knowledge on the topics covered in the lesson.

** Videos will be released based on the number of students actively participating in discussions and completing assignments. For the first video, a new video will be unlocked when 10+ students complete the previous assignment. For subsequent videos, the threshold will increase by 10 students for each video. For example:*

- *First Video: Unlocked when 10+ students complete the previous assignment.*
- *Second Video: Unlocked when 20+ students complete the previous assignment.*
- *Third Video: Unlocked when 30+ students complete the previous assignment.*
- *And so on...*

*** Students will have access to a dashboard showing their progress, engagement levels, and upcoming milestones, keeping them motivated to stay on track. A leaderboard will be maintained to recognize top performers and active participants.*

**** Outstanding contributions and projects will be highlighted by me, providing recognition and encouragement.*