**“Kyrgyz Headline Generation” Project Report**

**Done by: Aktan Raimbekov, Dastan David uulu, Elmirbek Orozobekov, Bayizbek Kutmanov, Argen Shamyrov, Elzar Zholdoshev**

**Main goal of the project:**

Our goal was to create a model, fine-tuned especially for Kyrgyz language, which can generate headlines for the long texts. The model generates short Headline using long text inserted.

**Team Contribution:**

Aktan Raimbekov - Main coder, code presenter

Dastan David uulu - Coder’s helper, Writed report, code presenter

Elmirbek Orozobekov - Database collector, Code presenter

Elzar Zholdozhev - Presentation design Creator, presentation presenter

Argen Shamyrov - Presentation text Creator, presentation presenter

Bayizbek Kutmanov - presentation presenter

**What we learned?**

We learned that Multilingual models based on transformers can be fine-tuned for not popular languages like kyrgyz.

Preprocessing and tokenizing are very important for model’s good performance

Also we learned that even small datasets if they are made with proper techniques can give good results.

Also it was our first time using “rouge” metrics. We found out it is really good metrics to compare generated and real answers. We learned about its stages of comparing texts.

**What can be improved:**

We could increase our dataset and improve the technique it’s made of so the model made better headlines.

Also we shortened texts by limiting them to 512 characters, maybe without limits model can better understand text and generate more accurate results. When we limit text ends suddenly without context and we think it can confuse model’s understanding of context.

Also would be better if we didn’t use multilingual models pretrained on much languages but used models specially trained for Kyrgyz language or if we trained it by ourselves so the model would understand the text better and generate better headlines.

Also we save not best result, but the last. It is because of some problems in trainer arguments. The number of epochs can also increase accuracy.