# Deberta Report

Maksim Zubok

2024-02-07

## Project Summary: ACLED Data Automatic Labelling

### Introduction

This document outlines the methodology and results of a project aimed at training a classifier for the automatic labelling of Armed Conflict Location & Event Data (ACLED) using hand-labelled observations.

### Model Selection

- The model chosen for this task was **DeBERTa Base**.
- No experiments were conducted with other Language Models (LMs) or different sized versions of DeBERTa.

### Dataset and Training

- The initial dataset comprised **4,405 observations**, which were split into training, validation, and test sets with **2,643**, **881**, and **881** observations respectively.
- The classifier was fine-tuned using the following hyperparameters:
  - Warm-up steps: **500**
  - Batch size: **8**
  - Weight decay: **0.01**
  - Learning rate: **1e-4** (to be confirmed by Maksim)
- The choice of batch size was influenced by cost constraints, allowing the training to run on a default Google Colab setup. Other parameters were selected based on default settings, and no hyperparameter tuning was conducted.

## Performance Metrics

The classifier's performance on the test set is summarized as follows:

| Label | Precision | Recall | F1 Score |
|---|---|---|---|
| Cultural | 87% | 100% | 93% |
| Economic | 84% | 89% | 86% |
| Environmental | 90% | 91% | 91% |
| Legal | 93% | 93% | 93% |
| Political | 99% | 95% | 97% |
| Social | 88% | 91% | 89% |
| War (anti) | 97% | 97% | 97% |
| War (pro) | 99% | 100% | 99% |

### Overall Performance

- **Micro Average**:
  - Precision: 95%
  - Recall: 95%
  - F1 Score: 95%
- **Macro Average**:
  - Precision: 92%
  - Recall: 95%
  - F1 Score: 93%

## Conclusion

The project demonstrates effective use of the DeBERTa Base model for the automatic classification of ACLED data, with significant precision and recall across a range of labels.