

## ERROR ANALYSIS

This section analyses the output produced by our best model for Task 1. In general, the model does not perform well on headlines containing extreme humor and is mainly affected by general world knowledge, sarcasm, and negative sentiments.

We report two error variations, classed as overestimation or underestimation. The table below presents an excerpt from our output dataset, which shows different error variations. The first headline with ID 11404 had the highest error score of difference 2.5 and the ground truth had the highest mean rating of 3 indicating the highest degree of funny. However, the model predicted a not funny score, mainly because the headline contained an extreme form of humor. Headline ID 11927, 1805 had an underestimation in the degree of humor which shows that the model lacks understanding of world knowledge.

ID	Original	Edited	Mean grade	prediction	error
11404	Hillary Clinton Supporters Filed a Complaint Against Bernie Sanders And Lost	Hillary Clinton Supporters Filed a Complaint Against themselves And Lost	3	0.5	-2.5
1805	RAF fighters intercept two Russian bombers near UK airspace   World news	RAF fighters intercept two Russian drunks near UK airspace   World news	3	1	-2
11927	Mike Pence Just Turned Over His Personal Emails, but Trump Is Still Focused on Hillary Clinton	Mike Pence Just Turned Over His Personal Emails, but Trump Is Still Focused on coloring	2.6	1	-1.6