**Ethical Considerations**

Group 3 MSU

**Write a one-paragraph story describing a fictional person who was positively affected by**

**a model trained with these data.**

Andrew runs a small-town movie theater. Every week he rotates which movies he plays based on what the people in the town want to watch, with our model he has access to a wild range of movies and he is able to get the most recommend movie to play based on what the people want to watch .

**Write a one-paragraph story describing a fictional person who was negatively affected by**

**a model trained with these data.**

Zu tang is a movie enthusiast who loves to watch movies, but he has trouble understanding English so he manly watches movies in mandarin. Zu tang is looking for new movies to watch. Zu tang was negatively impacted by our model because all the movies our model recommend are English oriented movies so people who don’t speak English cant use it.

**Describe at least two sources of bias the particular model in your story could have?**

One bias could be that there are no recent movies after July 2017 in the dataset, so the dataset is not up to data with the latest movies out. So, anyone looking for movies after July 2017 will not get a recommendation. Another bias could be that 70% of the movies in the dataset is in English and the other 30% percent are in different languages although its diverse it would be more diverse.

**Describe at least one way we could modify the model to mitigate this bias?**

One way we could mitigate bias in our model could be using other library’s apart from TF-IDF we could have used cosine similarity on latent semantic analysis (LSA/LSI) vectors which works better than TF-IDF or we could have used genism which runs faster and could have provided us with a better accuracy score.

**Describe at least one way we could modify the dataset to mitigate this bias?**

We could increase the amount of data we are working with so the dataset incorporates more recent movies and the movies in the dataset are as diverse as possible.

Describe at least one way we could modify the context surrounding the model to

mitigate this bias?

We could make this bias less severe by simply adding movies that are not strictly in the English language and other movies that have recent release dates to add more diversity to the dataset.