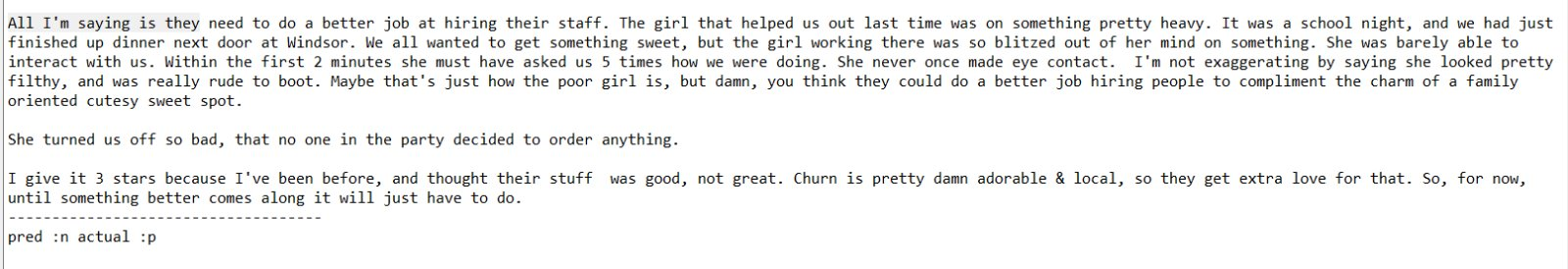
**BASELINE**

The two examples below are seen when the data was balanced and imbalanced.

After the data is balanced:



When we balanced the data, we got in total 749 rows for each star rating. It predicted it as negative. We think that due to the data is too small. However, if the data were too large, it would have predicted it right.

We were using bigram in this case, and we have spotted a few examples, that we think caused the system to predicted it negative even though it should be considered positive.

The examples:

**She never** once made eye contact.

I am **not exaggerating** by saying …….

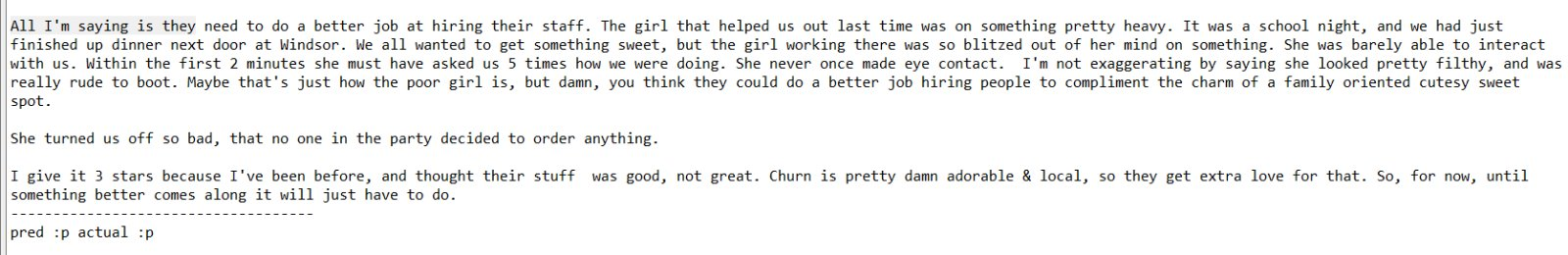
Just how the **poor girl** is

She turned us off **so bad**

Their stuff was good, **not great**.

We believe these examples made the system to predicted as negative.

When the data is not balanced:



When we left the data imbalanced, and most of our reviews are positive. It predicted it as positive regrading too many negative feedbacks in this review. We think that due to data is bias in this case.

These some examples mix of some negatives and positives comments:

They need to do a **better job** at hiring their stuff.

Wanted to get **something sweet**

She looked **pretty filthy**.

To compliment the **charm of** a family………

The girl that **helped us** out last time…….

**She never** once made eye contact.

I am **not exaggerating** by saying …….

Just how the **poor girl** is

She turned us off **so bad**

Their stuff was good, **not great**.

Churn is **pretty damn adorable**……

They get **extra love** for that.

Until **something better** comes along………

With imbalanced data it was predicted as positive, and it was expected, but we think that happened because the data was bias. A few of these positive bigrams showed up and with the help of biases. It got it right.

Finally, we need a huge dataset to robust our systems to overcome biases and to give our system more chances to learn better. We got a lot of reviews were predicted as positive-negative or negative-positive because either the data is bias or because there are no enough reviews to train our data on. Maybe with a huge dataset we will avoid these issues, not completely, but at least reduces them.