Improvements in model:

I have not implemented this feature, just as my utter focus is on making devOps operational with near to zero accuracy on models(in fact we are getting around 73%). Once we have pipeline in place, we can always have developers working on accuracy and ideas.

* **Outliers:** Can build outliers detection models to work on input data and find outliers percentage for us.

Outliers makes models robust and assist it to survive test data to much better accuracy that model without outliers. The thing is one should know, how much and what kind of outliers support us.

Also, as we are using SVM, it has a parameter called C which is the vital parameter focussing on boundary slopes separating the classes.

If C is very high, it is very sensitive to outliers and we should avoid using outliers for this case.

Thus, we develop a script which automates the process of selecting parameter C, based on the percentage of outliers out input data have. Techniques: point outliers: [isolation forest, z\_score]

Contextual outliers: [Tensorboard plotting & similarity & punctuation filters]

* **Feature Distribution stats for landing page lookup:** Currently, if we count the leads for each landing pages, the distribution is quite uneven and this we can hint the marketers to target and check the problem with those landing pages.

These can either end up in winding up that source (landing page) or can do a consolidated survey on what potential factors does one need to eliminate/maintain to bring that up.

* **Explainable AI:** This is the perfect example to build explainable AI. What it means is, once we train our model, we can predict it using data which includes landing\_page\_id and origin.

Once it has predicted, if correct, we can explain using the explanation we generated from our training data.

Extract Explanation:

We can group all data points by landing page id and further by source, which gives us some statistics to extract sentences from it. Also if we extrapolate and consolidate it further with win/lose leads statistics, we can explain it more on which landing\_page\_id goes well with which source and thus can explain prediction.

This is very naïve, but effective way to build explainable AI from scratch, since dataset is small and once we grow, we can get more insights in growing our explanation too.