**Problems encountered while running svm model.**

* During a trial run of the SVM model using certain features (like user\_awardee\_karma, user\_awarder\_karma, user\_link\_karma user\_comment\_karma, comment\_id, self\_text\_vader\_score, post\_title\_vader\_score, combined\_text\_vader\_score) all performance metrics such as precision, f1 score, and recall were reported as zero.
* Upon investigation, I discovered that this could be attributed to the following factors:
* Imbalanced dataset: If the dataset has a very low proportion of positive samples, the SVM model may struggle to learn the patterns and end up not predicting any positive samples correctly.
* Overfitting: The SVM model may have learned the training data too well and is not generalizing well to the test set.
* Inappropriate hyperparameters: The default hyperparameters of the SVM model may not be optimal for your dataset, and you may need to tune them to improve the performance.
* Further, I came to know that our dataset is highly imbalanced, with a much larger number of non-controversial (0) samples compared to controversial (1) samples.

This imbalance can cause issues with the SVM model, as it may struggle to learn the patterns in the minority class (controversial samples) and end up biasing the predictions towards the majority class.

To address this imbalance, I can try the following:

* **Oversampling the minority class**: Use techniques like SMOTE (Synthetic Minority Over-sampling Technique) to generate synthetic samples for the minority class, effectively increasing the representation of the controversial samples in the training data.
* **Undersampling the majority class**: Randomly remove some of the non-controversial samples to balance the class distribution.
* **Class weighting**: Assign higher weights to the minority class (controversial samples) during the SVM training process, to make the model more sensitive to the minority class.
* I attempted to address issue by using SMOTE to oversample the minority class. However, the model kept running and I had to interrupt the kernel.
* I then tried using class weighting, but it resulted in all zeros.
* Additionally, when I attempted a combination of oversampling and undersampling, this also kept running.
* I will run the same models with the final features and choose the model that gives the best results to proceed further.