

# The Data Chronicles:

Team Members: Anshuman Asauliya, Ayush Kumar, Shubham Garg

Public Leaderboard score: 24.67768

Private score: 25.09406

## **ALGORITHMS AND THOUGHT PROCESS USED:**

Encoding of data using splitting of created date, due date in days, months and years.

Created three columns DUE, INVOICE, CREATED converting days, months and years into number of days.

Created two columns due\_diff and invoice\_diff, which contained the difference between due and invoice dates and invoice and created dates respectively.

## **INSIGHTS INTO THE DATASET PROVIDED:**

Vendor name column was dropped as it was a string and did not prove to be much significant for the training data set.

Product description column was dropped because more than 85% values were missing.

Number of days until payment depends on the number of days between creation of invoice and due date, hence we have created separate columns due\_diff and invoice\_diff.

## **TECHNIQUES USED TO PROCESS THE DATASET:**

We tested the model results on XGBRegressor and StackingRegressor, which did not prove to be much accurate for the given data set.

We used gradient boost regressor model with customised parameters to train the model.