**CASSANDRA**

TEAM NAME: DATASIGHTS

TEAM MEMBERS:

VEDANTH POWAR

SHASHI KANT

SACHIN SHAW

**First Submission:**

Here, I prepared a scratch data set consisting of:

1. Amount float64
2. Settled float64
3. Outstanding float64
4. Created\_Time int64
5. Days\_to\_Pay int64
6. Created\_Day int64
7. Created\_Month int64
8. Created\_Year int64
9. Invoice\_Day int64
10. Invoice\_Month int64
11. Invoice\_Year int64
12. Due\_Month int64
13. Due\_Day int64
14. Due\_Year int64
15. Number\_of\_Days\_until\_Payment int64

**Algorithm:**

Gradient Boosting Regressor, and Score = 26.93958

**Thought process:**

We just broke the invoice, due, created date columns in useful information like:

1. Created Time
2. Month
3. Date of Month

We included these to give detailed event Review.

**Techniques:**

1. Strp() to convert string date to datetime object.
2. Split() to separate dates, months, year
3. To\_dict(), map() ==>> for mapping Year Data
4. And, other encoding Techniques.

**Intermediate Submission(18th):**

Here, we add more features to feed our model more accurate details of event:

Description int64

Vendor\_Name int64

Amount float64

Settled float64

Outstanding float64

Created\_Time int64

Days\_to\_Pay int64

Created\_Day int64

Created\_Month int64

Created\_Year int64

Invoice\_Day int64

Invoice\_Month int64

Invoice\_Year int64

Due\_Month int64

Due\_Day int64

Due\_Year int64

Invoice\_Week\_Day int64

Due\_Week\_Day int64

Created\_Week\_Day int64

Diff\_Invoice\_Creat int64

Number\_of\_Days\_until\_Payment int64

**Algorithm:**

Gradient Boosting Regressor with **n\_estimator = 550**, and Score = **24.55442**

**Thought process:**

Here we increased the features, giving more precise info of the event which improved our score.

Here we added:

1. Invoice\_Week\_Day
2. Due\_Week\_Day
3. Created\_Week\_Day
4. Diff\_Invoice\_Creat
5. Vendor\_Name
6. Description

**Week Day**: If holiday, vendor may be lenient on that week day.

**Vendor\_Name**: It gave a clear picture of Vendor Behaviour.

**Description**: It increased accuracies of some data points.

**Techniques:**

1. We concated the vendor from test and train data and made a combined map, which we used to give id to the vendor names.
2. Similarly, for Description.

**Final Submission(26th):**

Here, We modified some features:

Description int64

Vendor\_Name int64

Amount float64

Settled float64

Outstanding float64

Created\_Time int64

Days\_to\_Pay int64

Created\_Day int64

Created\_Month int64

Created\_Year int64

Invoice\_Day int64

Invoice\_Month int64

Invoice\_Year int64

Due\_Month int64

Due\_Day int64

Due\_Year int64

Invoice\_Week\_Day int64

Due\_Week\_Day int64

Created\_Week\_Day int64

Diff\_Invoice\_Creat int64

Number\_of\_Days\_until\_Payment int64

**Algorithm:**

Gradient Boosting Regressor with **n\_estimator = 1000**, and Score = **23.97915**

**Thought process:**

Here, we just wanted to increase the training iteration, which eventually improved the score.

And other thought process is added as **comments in the code for readability**.