Problem #1 Identify the model for Inter Company Transaction

Since it is a classification problem, I would go for algorithms like LogisticRegression / DesitionTree / RandomForest / Naive Bayes / SVM

Reason for classification algorithms : The Y vaue can be either yes / No

LogisticRegression:

LR is more sensitive to outliers

DesitionTree :

Decision trees implicitly perform variable screening or feature selection

Decision trees require relatively little effort from users for data preparation

Nonlinear relationships between parameters do not affect tree performance

Easy to interpret and explain to executives

Easy to interpret and make for straightforward visualizations.

The internal workings are capable of being observed and thus make it possible to reproduce work.

Can handle both numerical and categorical data.

Perform well on large datasets

Are extremely fast

RandomForest -

Advantges:

It doesnt overfit

One way Random Forests reduce variance is by training on different samples of the data

Random Forests require almost no input preparation.

Random Forests perform implicit feature selection

Random Forests are very quick to train

Random Forests are pretty tough to beat.

It’s really hard to build a bad Random Forest

Drawbacks:

The main drawback of Random Forests is the model size. You could easily end up with a forest that takes hundreds of megabytes of memory and is slow to evaluate.

Another point that some might find a concern is that random forest models are black boxes that are very hard to interpret.

Naive Bayes -

It requires less model training time

Naive Bayes classifier will converge quicker than discriminative models like logistic regression(so less training data)

Naive Bayes is that it’s a good algorithm for working with text classification

SVM

SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces

Problem 2: completeness and accuracy of given INTCO transaction

Problem #2 Identify the model for Intercompany Matching

Issue : Classsify if the trasaction needs booking or not

Model : Classification algorithm - LogisticRegression

Why : Need to review with Jothi to see if it is possible

Problem #3 Identify the model for Intercompany Booking

Per my knowledge, Booking cant be done using hte models

Problem #4 Identify the model for Intercompany Elimination

Issue : Identify transactions to eliminate intercompany transactions

Model : Same as prob #1 - Classification algorithm - LogisticRegression

Why : Intercompany trasactions need to be identified for elimination. Some what we need to identify is if it is a intercompany transaction or not. So LogisticRegression will work better

Problem #5 Create a consolidated financial Report

Per my knowledge, consolidated financial Report cant be done using hte models