**JetTrack Summary**

In this data research, we are trying to predict the potential deal status associated with two companies derived from one flight record.

**Constraints:**

During the research process, we have found the following constraint:

No specific target companies. We can use the aircraft\_id and the ownership table to map back which company is flying. However, there are tons of tracked companies around the arrival airport, so we cannot determine which one is the target company.

**Solutions:**

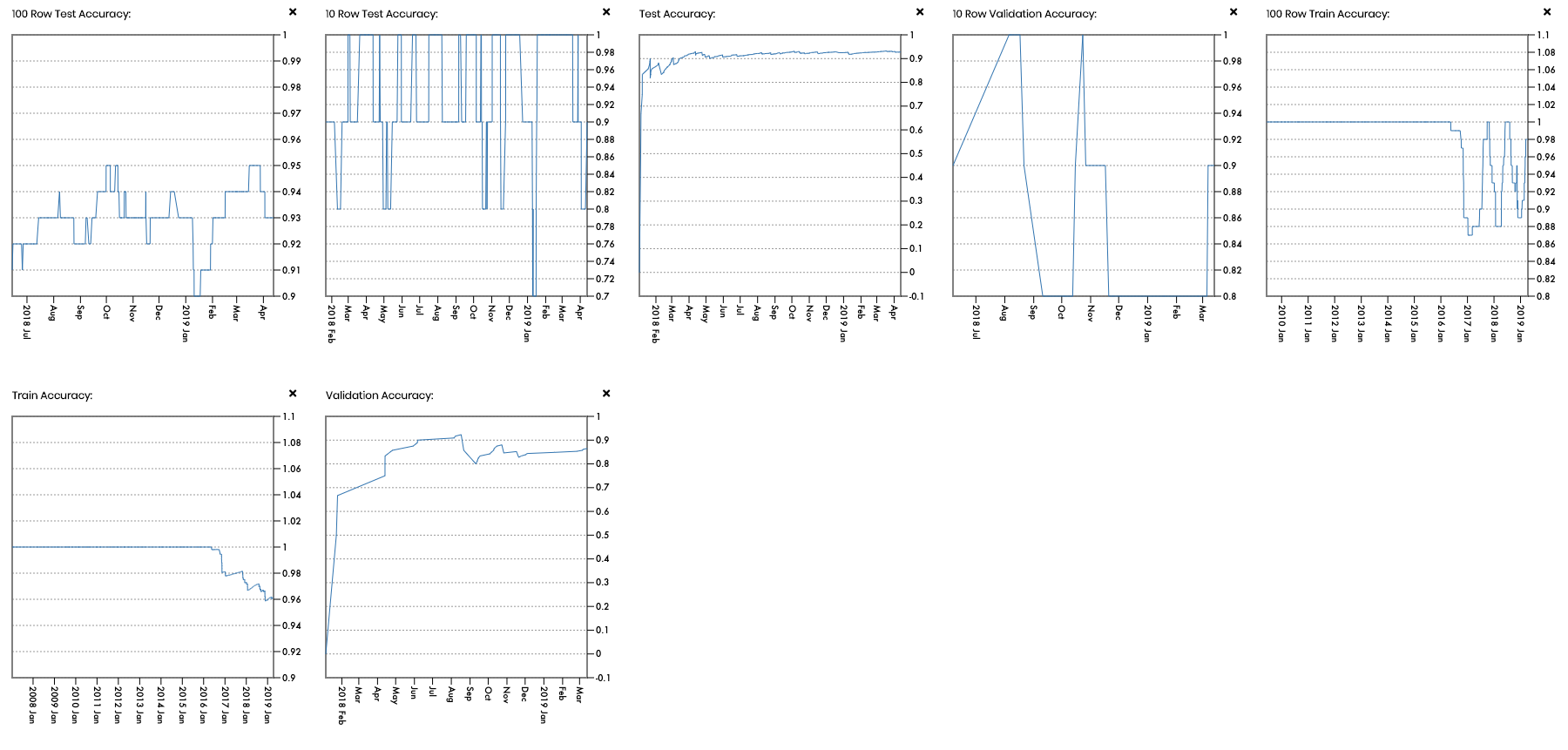
The ideal solution is to use a radius algorithm given the arrival airport coordinates and every company's coordinates. If the distance between the arrival airport and company is shorter than a defined radius (let's say 200 miles), then this company is included in a potential company list for one certain flight record.

After the detection, each flight record is expanded to multiple records equal to the number of companies in the list.

**Concerns:**

By doing so, we solve the issue of no target companies. However, another concern is raised, one where we have a large data file with a very unbalanced target. We expanded the 4 tickers’ flight records with the most transaction records in the transaction file. The total flight records of these 4 tickers is 6,212. After the expansion, the dataset contained 6.8 million records. Such a huge data file would take a long time to run. Additionally, among those 6.8 million records, only 186 of them have status other than “Nothing.” Therefore, the status is very unbalanced.

**Current Approach:**

Alternatively, we look at all the companies in the transaction file, and based on the target company, and coordinates, we filter out irrelevant flight records. Based on the evaluation charts in XYME, the model can predict the transaction status pretty well.  


**Future Improvement:**

1. In order to reduce the size of expanded flight records, we can combine JetTrack’s flight data with Accern's news analytics data. When a rumor is detected in Accern's data, it can be used as an indicator to look at JetTrack's records. All other records can be dismissed until a new signal appears.
2. Using machine learning techniques to determine the appropriate radius as we detect the target companies around the arrival airport.