Overview

In this practical, we were asked to investigate a real-world dataset involving flight information within the US and its territories and create a model using machine learning (ML) techniques to predict whether a flight will suffer from a disruption. This was done following a structure based on a subset of ML project structure. This consisted of exploring the data to learn about potential patterns that might affect disruption, manipulating the data to a format for the various machine learning models to train on, finding ways to improve the model, and then evaluating and critiquing the model on data it has never seen before.

Attribute Decisions

After exploring my chosen variables, I decided to not include the day of the week and expected arrival time block as attributes in my model. This was because after visualising the data, I could not see a clear correlation between it and disruption. On the other hand, the expected arrival time block was highly correlated linearly to the departure time block, thus adding both would be redundant for the model.

Training Performance and Insights

For the baseline model, I chose one that performed slightly worse but was still comparable to another type of classifier and was significantly faster to train. After training that chosen model, I ended up having an accuracy of 75%. However, this was not very well representative of the business aims as the model was heavily biasing predicting that a flight was not disrupted as 78% of the disrupted flights were predicted to be not disrupted. This is probably due to the imbalance of data favouring the non-disrupted class (As such I decided that balanced accuracy would be a better metric than accuracy, with this base model having 55%). With my interpretation, I believe that it is more important to predict that a flight will be disrupted than not disrupted. This is because in most cases, people are going to assume that their flight will not suffer any type of disruption, so this information is mostly not needed. It would therefore be more useful for potential users to predict whether their flight would be disrupted so that they could potentially account for longer travel. Even if a non-disrupted flight is predicted to be disrupted, I believe for users to find this out on the day would not cause any negative effects whereas if a user were to find that a flight predicted to be not disrupted was disrupted, their trust in using the model would decrease.

In an attempt to combat this imbalance, I tried to increase the weight of the disrupted class to influence the classification during training.

Another insight gained through training is that the region in which a flight takes off is not important to the model, thus it was removed as an attribute when fine-tuning the model.

Final Model

As a final model, despite the overall accuracy being lower (62%), balanced accuracy improved to 62%. Recall also improved from 21% to 62% but with a slight drop in precision from 31% to 29%. This gives a final F1-Score of 39%, 14% better than the initial model. This model as a whole fits the business objective better as 62% of the disrupted flights are now being correctly predicted in comparison to the initial solution of 22%. However, this model is limited as only ⅖ are still being incorrectly assigned, meaning there is a large room for improvement. This would probably require choosing a different set of attributes and some more trial and error.

Issues Encounter

When attempting to use the huge data set, the lab machine crashed when trying to visualise the data. Therefore, the large data set was used throughout the process. This means that there is a chance that the data used may not be fully representative of the actual data model which in turn could have influenced how the model trains.