Potential explanation: there is a column that encoders or a combination that gives you the exactly

One hot encoder only works with a specific classification

Classifier

Logistic regression – not label encoder ready

Gradient boosted tree is okay for label encoder ready

Pick a small dataset and do this experient and use labelencoder

Just because you’re walking forward in time doesn’t mean that you have unique methods

Use drop\_duplicates in test set.

Number of splits use the 7 years.

Usually happens in Kaggle (google data leak)

Order of importance

1. Look for data leak
2. Remove duplicates
3. Change time period
4. Try one-hot-label encoder (iris dataset)

When you get a high accuracy vs high accuracy through iteration

Highway-railroad crossing outliers

1. Prescription – What would you do about the situation? Now give a little more about that, think about what can be done? Investigate deeper on the data based on the data.

Location -