Note

* Local CV (cv and lb consistent mean reliable. Checks if the model predicts well into the distant future rather than the near future)

<https://www.kaggle.com/c/ieee-fraud-detection/discussion/107728>

* + Time Series
  + Group k fold (6 fold)
* Feature Selection
  + Correlation
  + Permutation importance for solo features
  + Recursive feature elimination for block of features

<https://www.kaggle.com/nroman/recursive-feature-elimination>

* + PCA for groups of identical features (V)
  + Some discussion

<https://www.kaggle.com/c/ieee-fraud-detection/discussion/102940#latest-620260>

* Deeper feature engineering:
  + Noise
  + Outliner
  + Normalization
  + Frequency encoding
  + Bool encoding
  + Timeblock frequency encoding
  + Aggregation
  + Some FE example:

<https://www.kaggle.com/nroman/eda-for-cis-fraud-detection>  
<https://www.kaggle.com/artgor/eda-and-models>  
<https://www.kaggle.com/kyakovlev/ieee-gb-2-make-amount-useful-again>  
[https://www.kaggle.com/felipemello/why-your-model-is-overfitting-not-making-progress](https://www.kaggle.com/felipemello/why-your-model-is-overfitting-not-making-progress(This)

* Model
  + lgbm
  + xgboost
  + catboost
  + NN
  + Useful kernels

<https://www.kaggle.com/timon88/lgbm-baseline-small-fe-no-blend>  
<https://www.kaggle.com/nroman/lgb-single-model-lb-0-9419>  
<https://www.kaggle.com/artgor/eda-and-models>

* Ensemble
* How all work together:

<https://www.kaggle.com/c/ieee-fraud-detection/discussion/107697>

* About the datasest:
  + Test&train split

<https://www.kaggle.com/c/ieee-fraud-detection/discussion/101040#latest-609949>

* + Information of card1~6

<https://www.kaggle.com/c/ieee-fraud-detection/discussion/107833#latest-621858>

* + Time of day

<https://www.kaggle.com/c/ieee-fraud-detection/discussion/100400#latest-616798>

* + D9 is the decimal part of D8. I think D8 is the days passed since last transaction and D9 is hours passed since it is the decimal part…
* Pipeline