Library: [Lantianzz/Scorecard-Bundle: A High-level Scorecard Modeling API | 评分卡建模尽在于此 (github.com)](https://github.com/Lantianzz/Scorecard-Bundle)

Example to follow: [Scorecard-Bundle/[Example] Training a Scorecard model using California Housing Dataset.ipynb at master · Lantianzz/Scorecard-Bundle (github.com)](https://github.com/Lantianzz/Scorecard-Bundle/blob/master/examples/%5bExample%5d%20Training%20a%20Scorecard%20model%20using%20California%20Housing%20Dataset.ipynb)

Variable/feature selection – [sklearn.feature\_selection.RFECV — scikit-learn 0.24.2 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFECV.html)

Binning - [sklearn.preprocessing.KBinsDiscretizer — scikit-learn 0.24.2 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.KBinsDiscretizer.html)

* Fine binning (slides 116-134)
  + Autobinning (slides 129-134)
  + Use ‘quantile’ as discretization strategy (supervised)
* Coarse binning (slides 136-145)
* Questions
  + Should fine binning be unsupervised (use ‘uniform’) and coarse binning be supervised ?

Creating a scorecard model (Section 6, slide 100+)

* Using

Building scorecard (section 9, slide 147+)

* Conceptual Approach to Scorecard Development
  + Train/test split
  + ‘roc\_auc’ scores
* Variable Selection
  + Auto variable selection
  + Manual variable selection
* Model Version History

7/1

Fine binning is done by ‘initial\_interivals’ param in ChiMerge call

Coarse binning is done by ‘max\_intervals’/’min\_intervals’ params

To do:

* Handle missing values → Assign average (weighted) score used for each feature

Slide 14:

* Allows for "Bagging" (Bootstrap Aggregation) of scorecards, a process in which many bootstrap replicas of a scorecard are fitted and then averaged.
* Allow problem variables to be de-emphasized by range restriction for a given set of variables while trading off overall model strength.
* Handles mixed variables (i.e., which take on continuous and categorical values).
* Handles missing values (i.e., no information) without having to drop the observation. Missing values are forced to take on the neutral predictive score weight.
* Provides automated and/or interactive fine and coarse binning of predictors, as required for developing Generalized Additive Models.