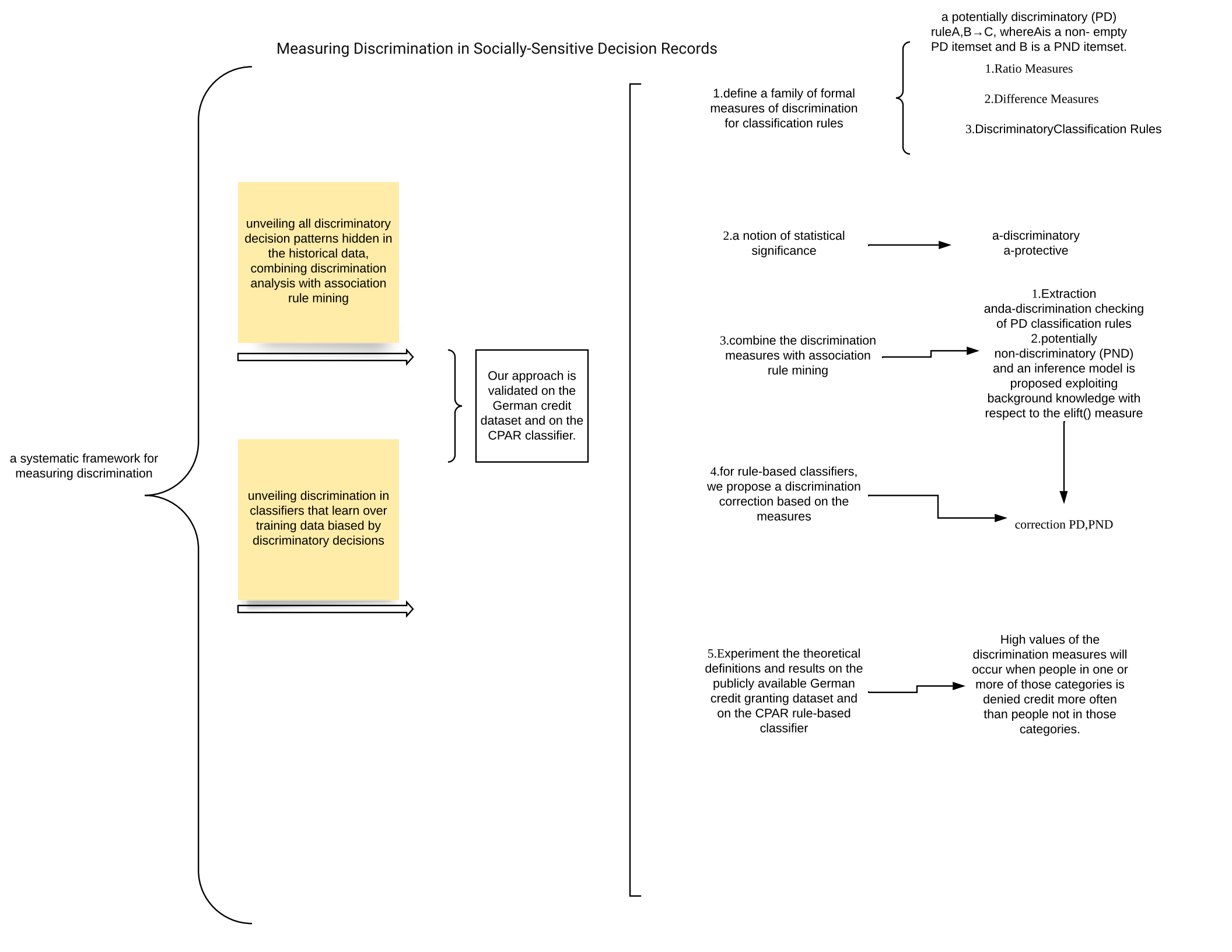
Measuring Discrimination in Socially-Sensitive Decision Records
a systematic framework for measuring discrimination
unveiling all discriminatory decision patterns hidden in the historical data, combining discrimination analysis with association rule mining
unveiling discrimination in classifiers that learn over training data biased by discriminatory decisions
Our approach is validated on the German credit dataset and on the CPAR classifier.
a potentially discriminatory (PD) ruleA,B→C, whereAis a non- empty PD itemset and B is a PND itemset.
1.define a family of formal measures of discrimination for classification rules
1.Ratio Measures
2.Difference Measures
3.DiscriminatoryClassification Rules
2.a notion of statistical significance
a-discriminatory a-protective
3.combine the discrimination measures with association rule mining
1.Extraction anda-discrimination checking of PD classification rules
2.potentially non-discriminatory (PND) and an inference model is proposed exploiting background knowledge with respect to the elift() measure
4.for rule-based classifiers, we propose a discrimination correction based on the measures
correction PD,PND
5.Experiment the theoretical definitions and results on the publicly available German credit granting dataset and on the CPAR rule-based classifier
High values of the discrimination measures will occur when people in one or more of those categories is denied credit more often than people not in those categories.