**Enabling Machine Learning Algorithms for Credit Scoring - Explainable Artificial Intelligence (XAI) methods**

<https://arxiv.org/ftp/arxiv/papers/2104/2104.06735.pdf>

**0. Abstract and Introduction**

|  |
| --- |
| * 이 논문에서는 credit scoring을 어떻게 분석하는지 보여준다. 즉 **다양한 predictive model들을 비교하고, advanced tree based model이 가장 잘 예측한다**는 것을 보여준다. * 또한, **advanced model을 해석**하고 credit risk practitioner들의 접근을 보다 쉽게 하는 기술을 통해 **이들을 어떻게 boost**할지 보여준다. * 이들은 모두 **Polish Credit Bureau로부터 얻은 large dataset**에 대해서 보여진다. |

**[Keywords] credit scoring, model, scoring, XAI**

**1. Algorithms for scorecard construction**

|  |
| --- |
| 이 논문에서는 [18]에 제시된 접근 방법을 이용하며, **borrower가 default인지 아닌지를 예측하는 다양한 모델을 비교**한다.   * 데이터는 로 나타낸다.   + 는 good/bad borrower를 나타내는 0 또는 1의 값인 **explanatory variable**이다.   + 는 **explanatory variable**들의 값을 의미한다. * 여기서는 [18]에 나타나 있는 알고리즘들 중 일부를 선택하여, **최근에 개발된 모델에 대한 추가적인 연구를 포함**한다.   여기서는 다음의 알고리즘을 비교한다.   * **Logistic regression** (borrower에 대한 default의 가능성을 모델링) * **Logistic regression with weight of evidence (WOE) transformation** 접근 방법   + WOE method는 transformed prediction과 explanatory variable 간의 비선형 dependence를 보여준다. * 다른 방법들은 현대의 머신러닝 알고리즘이다.   + **많은 decision tree 구축을 통한 Random Forest 접근 방법**     - 각 decision tree가 어떤 예측을 생성하고, 최종 예측은 이들 중에서 선택하며 특정한 값의 frequecy를 이용한다. * 또 다른 방법은 **gradient boosting (GBM)**과 **extreme gradient boosting (XGB)**이다.   Credit risk 모델링을 위해 현대의 방법을 사용하고 이들을 비교하는 것뿐만 아니라 **XAI tool을 이용하여 이들의 모델을 설명**하고, **복잡한 알고리즘의 결정에 숨은 로직을 이해하고 특정한 feature가 어떻게 예측에 영향을 미치는지를 논의**한다. |

**2. Experimental Results**

|  |
| --- |
| 이 논문에 이용된 dataset은 Polish Credit Bureau에서 얻어왔으며, **500만 개 이상의 고객 관찰에 대해 1729개의 variable**이 포함되어 있다. 또한 날짜 범위는 2017년 1월 10일부터 2019년 5월 30일까지이다.   * 6개월 default window를 가정하기 때문에, 모델링 표본을 2018년 11월 30일까지의 데이터로 제한했다. * 분석 모델의 품질을 평가하기 위해, **전체 dataset을 4개의 subset으로 나누고 이것을 training set**이라고 한다. * **Test set**은 accuracy와 stability를 평가하기 위한 것으로, **2018년 8월 31일부터 11월 30일까지의 데이터**로 구성되어 있다. 이것은 모델을 fitting하는 데 사용되지 않고, 나중에 effieicncy를 평가하기 위해 사용된다.   모델링 이전에, **모든 variable은 descriptive statistics라는 문맥에서 실험**되었으며, missing data는 linear model의 disruption을 방지하기 위해 평균으로 채웠으며, 일부 variable에 대해서는 dummy encoding을 실행했다.   * Predictor의 개수가 매우 많으므로, 현대의 머신러닝 알고리즘에서는 문제가 되지 않지만 **logistic regression과 같은 전통적인 알고리즘**에 대해서는 문제가 된다.   이 논문에서는 **Xgboost 접근 방법**에 기반한 방법을 사용했는데, 이때 **변수들은 subset으로 나누어지고, 이들은 unique value의 개수로 구분된다.**   * Division threshold는 300으로 설정되었으며, 이것은 변수들의 unique value의 개수의 중간값이다. * 이들 set에 대해서 **xgboost 모델**이 생성되고, 이것은 **81개의 변수에 대한 predictive power**를 나타낸다. * 선택된 변수에 대해 **Kolmogorov-Smirnov (K-S) statistic**을 계산하고, 0.1보다 작은 값은 제거된다. 따라서 **63개의 변수만 남는다**. 이 set을 이용하여 모델을 평가한다.   여기서 강조하려는 것은 **이들 중 선택된 34개의 변수**에 대해서는 Random Forest, gbm, xgboost뿐만 아니라 **Neural Network, Naïve Bayes, Regularised Logistic Regression과 같은 다른 모델로도 테스트**를 했다는 것이다. |

**2. Experimental Results (Cont.)**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 여기서는 **Gini measure**를 이용하며, **rejection은 Gini < 0.6**일 때이다.  위 그림 **(Figure 1)**의 각 항목은 다음과 같다.   |  |  | | --- | --- | | Random forest **(rf\_63\_RS)** | training에서 **1.0**이라는 높은 Gini value를 보였으며, test data set에서는 **0.76**을 보였다. | | GBM **(gbm\_63)** | 각각의 set에서 Gini value가 **0.68**이다. | | Logistic Regression with WOE transformation **(reg\_log\_63\_woe)** | Logistic model class에서 가장 좋은 모델이다. | | Pure Logistic regression model **(reg\_log\_63)** |  | | Xgboost **(xgb\_63\_RS)** |  |   또한, **out-of-sample dataset**에서의 모델 비교 결과는 다음과 같다.   * **녹색 점 (out-of-sample)**, **검은색 점 (test sample)** * **Training, test, out-of-sample에서의 Gini 값의 차이**는 random forest 알고리즘을 제외하고 모두 **0.001** 정도이다.   이것을 표로 나타내면 다음과 같다. |

**3. Explainability**

|  |
| --- |
| 이제 **XAI tool을 통해 모델에 대한 이해를 얻는다.**   * Variable name은 **safety reason** 때문에 self-explanatory하지 않지만, 그 결과를 확인할 수 있고 그 속성에 대해 논할 수 있다.   **[Global Interpretations]**  Global interpretation 방법에서, **개별적인 변수들의 모델에서의 일반적인 역할**에 대해 이해하고, **prediction에서의 평균적인 impact**에 대해서 논할 수 있다.  여기서 적용한 tool은 **PDP profiles**와 permutation에 기반한 universal method인 **permutation feature importance (PFI)**이다.   * 여기서는 **straightforward and easy to use**인 후자의 방법을 이용한다.   + 이 방법은 universal하고, 이 논문에서 고려되는 모든 모델에 적용할 수 있다.   **GBM model**의 문맥에서 그 결과는 다음 그림과 같다.     * **각 feature들을 여러 번 random shuffling**한 후에는 **정확도의 highest mean decrease**를 관찰할 수 있다. * 따라서 이를 통해 **특정 변수가 defaulter와 non-defaulter를 구분하는 데 가장 영향력 있는 변수**라는 것을 알 수 있다. |

**3. Explainability (Cont.)**

|  |
| --- |
| 여기서는 **gradient boosting model (GBM)에 대해 PFI**를 적용했으며, feature OAG\_7이 가장 중요한 것으로 나타났다.  **PFI를 서로 다른 모델에 사용**할 수 있고 **다양한 경우에 대해 조사**하여, **discrimination 과정에서 어떤 변수가 가장 중요한지**를 알 수 있다는 것이 중요하다.    위 그림의 PDP plot에서 **가장 중요한 변수에 대한 추가적인 insight**을 얻을 수 있다. **PDP profile**은 **주어진 모델에서 변수의 값들 및 평균 prediction 간의 dependency**를 보여준다.   * 여기서는 모든 모델의 예측이 **OAG\_7이라는 변수와 유사하게 변화**하며, 이 feature의 값이 클수록 average prediction의 값이 감소한다. * 이와 유사한 변화가 **classical (WOE, GLM), modern (GBM, XGB) 모델**에서 나타난다. * **각 모델에 대해 PDP profile이 어떻게 변화하는지 관찰**하여 **모든 모델이 서로 같은 relation을 capture하려고 한다**는 것을 알 수 있다. (faster decline -> stability) |

**3. Explainability (Cont.)**

|  |
| --- |
| **[Local methods]**  Local method들은 **특정 prediction의 결정에 영향을 미친 crucial factor**를 설명하는 것을 목적으로 한다.   * 단일 observation에 대한 **PDP profile의 local equivalent**는 **다른 변수가 모두 fix된 상태에서 분석 대상 변수가 변할 때 prediction value의 변화**를 보여준다.   아래 그림에서는 **특정 고객**에 대한 **선택된 변수**에 대한 CP plot을 보여준다.    예측 시각화에 도움이 되는 다른 유용한 도구는 **선택된 customer에 대해 prediction을 분해**하는 **BD plot**이다.   * 아래 그림에서는 final PD가 0.051이고, 분석된 sample에 대한 평균 prediction이 0.559라는 것을 알 수 있다. |