

Abstrakt

Jednou z najzákladnejších úloh kreditného skóringu je klasifikácia potencionálnych klientov žiadajúcich o úver na "dobrých" alebo "zlých", na základe pravdepodobnosti, že by neboli schopní splácať úver v prípade, že by im bol odsúhlasený. V tejto práci porovnávame často používanú logistickú regresiu, neuronové siete a ensemble modely založené na stromových metódach. Pri konštrukcii neuronových sietí používame nové metódy a poznatky z oblasti hlbokého učenia, zatiaľčo v prípade stromov používame populárne ensemble algoritmy bagging, boosting a náhodné lesy. Modely porovnávame na základe ROC AUC miery, ktorá by mala poskytnúť väčšiu informačnú hodnotu ako len samotná presnosť. Výsledky naznačujú malý alebo takmer žiadny rozdiel medzi modelmi. V najlepšom prípade, dosahujú neuronové siete, boosted ensemble modely a zložené ensemble modely len približne o 1% – 2% väčšiu ROC AUC hodnotu ako logistická regresia.

Kľúčová slova

kreditní skóring, neuronové siete, rozhodovací strom, bagging, boosting, náhodný les, ensemble, ROC krivka