

Aplikácie metód strojového učenia v oblasti zdravotného poistenia

Vedúci práce: RNDr. Ľubomír Antoni, PhD.

Konzultanti: RNDr. Peter Bugata , PhD.

RNDr. Dávid Hudák , PhD.

Úvod do problematiky

- Popis práce revízneho lekára
- Veľké zaťaženie v líniovej revízii
- Automatizácia revíznej činnosti:
 - Zníženie počtu výkazov
 - Zvýšenie konzistentnosti rozhodovania
- Pomocou metód strojového učenia

Ciele práce

- Oboznámenie sa so základmi strojového učenia
- Preštudovanie konkrétnych metód klasifikácie
- Aplikovanie daných metód na reálnom probléme
- Preskúmanie algoritmov založených na rozhodovacích stromoch

Strojové učenie

- Štúdium algoritmov a matematických modelov, ktoré progresívne vylepšujú svoje riešenia pre špecifické úlohy
- Učenie s učiteľom a bez učiteľa
- Binárna klasifikačná úloha

Popis datasetu

Počet pozorovaní (riadkov): 1 234 821

Počet atribútov (stĺpcov): 437 - 7758

Rozdelenie na tréningovú a testovaciu množinu: 70 / 30 , stratifikovane

	<i>Premenné</i>	<i>Typ</i>	<i>Hodnoty</i>
1	Zaradenie dávky	<i>Categorical</i>	46
2	Príznaky chýb	<i>Binary</i>	391
3	Diagnóza	<i>Categorical</i>	1 705
4	Kód druhu produktu	<i>Categorical</i>	11
5	Modifikovaný produkt	<i>Categorical</i>	5 604
6	Cena vypočítaná	<i>Numeric</i>	1
			7 758

Vyhodnotenie modelov

Uvažované atribúty: Zaradenie dávky, Príznaky chýb

	Metóda	Accuracy
1	Naive Bayes	77.03 %
2	Logistic Regression	79.80 %
3	kNN	83.17 %
4	SVM	83.92 %
5	Decision Tree	84.14 %
6	<i>Neural Network (dense)</i>	84.15 %
7	Random Forest	84.19 %
8	Gradient Boosted Trees (XG Boost)	84.19 %
x	Teor. max. accuracy	84.71 %

Kódovanie a stromové algoritmy

Model	Presnosť predikcie (% , testovacia množina)			
	LabelEnc	RandomLabelEnc avg	Standard Deviation	OneHotEnc
decision tree	91.49	91.06	0.036	91.74
random forest	92.08	91.80	0.022	92.17
gradient boosting	92.41	92.14	0.026	92.05

Štúdijná literatúra

- RASCHKA, S. Python Machine Learning, 1st Edition. Birmingham : 2015. ISBN 978-1-78355-513-0.
- ZAKI, Mohammed J. - MEIRA, Wagner Jr. Data Mining and Analysis: Fundamental Concepts and Algorithms. New York : 2014. ISBN 978-0-521-76633-3.
- GUO, C. - BERKHAM, F. Entity Embeddings of Categorical Variables. [online] 22 Apr. 2016. Dostupné na: <<https://arxiv.org/abs/1604.06737>>

Ďakujem za pozornosť