

Mercari - Price Suggestion Challenge

Bojana Đerić, Ivan Emanuel Pavlov,
Tatjana Ramljak, Lara Rajković

30. travnja 2019.

Opis problema i cilj istraživanja

- Mercari - japanska aplikacija za prodaju i kupnju proizvoda.
- Cilj je napraviti što bolji model koji predlaže cijenu proizvoda.
- Natjecanje na Kaggleu
- Mjera evaluacije je *Root Mean Squared Logarithmic Error (RMSLE)*

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i + 1) - \log(a_i + 1))^2}$$

pri čemu su

n - broj podataka u skupu,

p_i - predviđene cijene,

a_i - stvarne cijene.

Jedan *listing* sadrži:

- name
- item_condition_id
- category_name
- brand_name
- shipping
- category_name
- item_description
- ciljnu varijablu price.

Problematične su nedostajuće vrijednosti i nepraktičnost tekstualnih podataka.

Skup za treniranje ima 1 482 535 različitih *listinga*, dok ih skup za testiranje ima 693 359.

Pregled dosadašnjih istraživanja

- Neki od korištenih algoritama su višeslojni perceptron (*MLP*), konvolucijske neuronske mreže (*CNN*), *Light Gradient Boosting*, *Random forest*, *Ridge* regresija...
- Korišteni su ansambli tih modela.
- Najmanji, pobjednički *RMSLE* jednak 0.37758.

- Fokus na *feature engineeringu* - *TF-IDF* za tekst, *PCA* i *MCA* za smanjenje dimenzionalnosti.
- Modeli: linearna regresija s normalizacijom, *random forest*.
- Metode dubokog učenja: višeslojni perceptron (*MLP*), konvolucijske neuronske mreže (*CNN*), povratne neuronske mreže (*RNN*).
- paketi: wordcloud, textblob, sklearn, tensorflow, keras...