

- One-hot kodiranje: kategorije, meseci, dnevi v tednu in ure
- Nove značilke: število slik, dolžina članka in število odstavkov
- Bag of words: naslov, podkategorije, uvod, ključne besede in GPT ključne besede
- Bag of Words → TF-IDF → najbolših 250 besed (Pearsonov koeficient korelacije)
- Omejitev maksimalnega števila komentarjev v učnih podatkih na 90

topic_zabava-in-slog	topic_znanost-in-tehnologija	n_paragraphs	length	n_pictures
				
false	false	4	1350	1
false	false	8	3002	1
false	false	5	2085	1
false	false	3	747	1
false	false	6	2067	2
false	false	33	6232	5

# VREDNOTENJE MODELA

10-kratno prečno preverjanje

Linearna regresija:

- R-kvadrat: 0.264
- Spearmanov koeficient korelacije: 0.529

Ridge regresija (alfa: 0,01):

- R-kvadrat: 0.127
- Spearmanov koeficient korelacije: 0.365

# RAZLAGA MODELA

Linearna regresija:

energija, liga nba, nba, domena, zelenski, ministrica, golob, košarka, janša, ukrajina, skok, zajc, roglič

Ridge regresija:

število odstavkov, kultura, svet, šport, ob 7, slovenija, zabava in slog, ob 8, gospodarstvo, število slik, ob 14, ob 9, okolje