

## ПАЈТОНСКИ ЈУРИШ НИЗ „НОРВЕШКА ШУМА“

---

Марко Димовски <sup>1</sup>

<sup>1</sup> ЕВН Македонија, Скопје

e-mail: mdimovski@gmail.com

„Норвешка шума“ е едно од најдобрите дела на јапонскиот писател Харуки Мураками. Каква е поврзаноста меѓу математиката и книгите? Со примена на низа статистички модели во програмскиот софтвер Python, ќе биде направена обработка на зборовите користени во книгата, нивната просечна должина, просечната должина на речениците (бројот на зборовите од кои е составена), а ќе биде извршено и математичко класирање на содржините од страниците на книгата. Како зборовите од една книга можат да преминат во нумерички променливи кои понатаму можат да се користат како предиктори во разни модели на регресија и класификација? Користејќи дел од техниките за „обработка на природните јазици“ (natural language processing) во Python, ќе направиме модел за предвидување на оценката која читателот ја дава како рејтинг за книгата, користејќи множество од критички осврти преземени од [www.goodreads.com](http://www.goodreads.com).