|  |  |  |
| --- | --- | --- |
| **Univerzita Hradec Králové**  **Fakulta informatiky a managementu**  **Katedra informatiky a kvantitativních metod** | | |
| **Strojové učení na platformě Apache Spark**  Bakalářská práce | | |
| Autor: František, Hylmar  Studijní obor: Aplikovaná informatika | | |
| Vedoucí práce: prof. RNDr. PhDr. Antonín Slabý, CSc.  Odborný konzultant: Ing. Ondřej Klapka | | |
| Hradec Králové | březen 2020 | |
| Prohlášení:  Prohlašuji, že jsem bakalářskou práci zpracoval samostatně a s použitím uvedené literatury. | | | |
| V Hradci Králové dne 10.4.2020 | | František Hylmar | |
| Poděkování:  Děkuji vedoucímu bakalářské práce prof. RNDr. PhDr. Antonínu Slabému, CSc. za metodické vedení práce a neocenitelnou pomoc při její tvorbě. | | | |

Anotace

Bakalářská práce se zaměřuje na aplikaci postupů strojového učení na velké objemy dat pomocí platformy Apache Spark. V práci jsou popsány metody doporučování produktů pomocí algoritmů pro kolaborativní filtrování. Vybraný algoritmus je implementovaný v programovacím jazyce Scala a nasazen na virtuální Hadoop clustr. Pro dataset obsahující implicitní data o poslechu hudebních interpretů je na základě implementovaného algoritmu vytrénován a optimalizován příslušný model. Přesnost predikcí modelu je následně ohodnocena jak strojově, tak i manuálně.

Annotation

Title: Machine Learning with Apache Spark

This bachelor thesis focuses on the application of machine learning procedures to large volumes of data using the Apache Spark platform. This work describes methods of product recommendation using collaborative filtering algorithms. Selected algorithm is implemented in Scala programming language and deployed on virtual Hadoop cluster. For a dataset containing implicit music data, the model is trained and optimized based on the implemented algorithm. The accuracy of model predictions is then evaluated both automatically and manually.