EuroSciCon

8th Edition of International Conference on

Big Data &
Data Science

March 04-05, 2019
Barcelona, Spain

Machine learning with spark

Plyushchenko Andrey N

Eastwind, Russia

Spark is the one of the most popular tools for effective Big Data manipulation with high-level languages such as Python, Scala, etc. PySpark is a Python-library for spark using. Although Spark includes a library of machine learning algorithms, the most popular local machine libraries such as SKLearn, XGBoost, etc., are more flexible and give the best results. We describe some techniques, which allow fitting standard algorithms and predicting values for distributed data.

Recent Publications

1. A N Plyushchenko and A M Shur (2011) Almost overlap-free words and the word problem for the free Burnside semigroup satisfying $x^2=x^3$. Internat. J. Algebra Comput. 21:973-1006.

Biography

Plyushchenko Andrey N has completed his PhD at Ural Federal University. He has completed School of Data Analysis at Yandex. Currently, he is a Head of Data Science Department in Eastwind, Software Development Company. He works with projects related to machine learning, Big Data, Data Analysis, etc. He has published about eight papers in reputed journals.

a.plyushchenko@eastwind.ru