

# Challenges and Opportunities in Analytics of Big Data

**Zlatinka Kovacheva**

Department of Mathematics and Applied Sciences, Middle East College

P.B. No. 79, Al Rusayl, PC: 124, Muscat (Oman)

Email : [zlatinka@mec.edu.om](mailto:zlatinka@mec.edu.om)

## ABSTRACT

Big data analytics uses advanced analytic techniques against very large, diverse data sets that include different types such as structured/unstructured, streaming/batch, and different sizes - from terabytes to zeta bytes. Big Data has one or more of the following characteristics – high volume, high velocity, high variety, high veracity, high value, and high vitality.

The variety and complexity of the data nowadays require diversity of methods applied for data analytics. Every method for big data analysis has its advantages and drawbacks. The big number of the methods ensure their application according to the specifics of the data and its required analysis. The purpose of this paper is to analyse and outline them in order to support data analysts to select and apply the most suitable techniques for their specific data sets and analytical goals.

The KEY technologies for Big Data Analytics are discussed. Statistical analysis vs. machine learning approach is presented. The main drawbacks of the classical statistical techniques are emphasized. The 3-staged process of data mining and its main methods have been discussed. The focus lays at the main advantages of neural networks in data analysis process.

## Предизвикателства и възможности в анализа на големите данни

**Златинка Ковачева**

Департамент по математика и приложни науки, Близкоизточен колеж P. B. № 79, Al Rusayl, PC: 124, Маскат (Оман)

Имейл: [zlatinka@mec.edu.om](mailto:zlatinka@mec.edu.om)

## РЕЗЮМЕ

Анализът на големи данни използва усъвършенствани аналитични методи за обработка на много големи, разнообразни набори от данни, които включват различни типове (като структурирани / неструктурирани, поточни / партидни), с различни размери - от терабайти до зета байтове. Големите данни имат една или повече от следните характеристики - голям обем, висока скорост, голямо разнообразие, висока правдоподобност, висока стойност и висока жизненост.

Разнообразието и сложността на данните в наши дни изискват разнообразие от методи, използвани за техния анализ. Всеки метод за анализ на големи данни има своите предимства и недостатъци. Големият брой методи осигурява тяхното приложение според спецификата на данните и необходимия анализ. Целта на настоящата статия е да ги посочи и сравни, за да подпомогне анализаторите на данни да избират и прилагат най-подходящите средства за техните специфични набори от данни и аналитични цели. Обсъждат се ключовите технологии за анализ на големи данни. Сравнен е подходът на статистическия анализ с машинното самообучение. Основните недостатъци на класическите статистически методи са очертани. Обсъден е триетапният процес на извличане на знания и основните му методи. Акцентира се върху основните предимства на невронните мрежи в процеса на анализ на данни.