

According to the article, Big Data is a relatively new concept and still there is no complete consent about its strict definition. However, there are different conditions which, if met, may denote that a particular data source can be described as Big Data. These conditions can be briefly described as “Three V’s”, which means Volume, Variety and Velocity. Volume refers to the size of collected data, often exceeding terabyte or petabyte. This also depends on the type of data, for example videos, by itself, take up more space than tabular data. Variety refers to different possible structure of a dataset. For example structured data, that is data prepared properly to be processed by computers like spreadsheets and databases, constitutes only 5% of all collected data. Most of the existing data is unstructured like videos and pictures, which is relatively cumbersome for automated processing. Velocity refers to the speed at which data is generated by a data source like a continuous live stream of mobile application data describing customer behavior and past buying patterns. The other criteria for Big Data classification are: Veracity which describes unreliability of the data like opinions in social media, Variability and Complexity which describe variations in the data flow rates and variations of sources generating data, Value which describes data which has a low value in its raw form but may gain very high value when properly processed and analyzed. Big Data analytics refers to modeling, analysis and interpretation of collected data. It can be divided into various categories depending on the type of the data: text analytics, audio analytics, video analytics, social media analytics and predictive analytics. Text analytics refers to extracting information from textual data and comprises of the following techniques: information extraction, text summarization, question answering and sentiment analysis. Audio analytics refers to extracting information from audio data for example speech recordings and comprises of transcript-based approach and phonetic-based approach. Video analytics refers to analyzing and extracting information from video streams and can be performed in server-based architecture or edge-based architecture. The main goal of social media analysis is to gather useful information from social media channels. It comprises of user-generated content analysis like content-based analysis and structure-based analysis and social network analysis like community detection, social influence analysis and link prediction. Predictive analytics is used for predicting future outcome based on collected data and the techniques used here can be divided into different groups based on various factors, for example: regression techniques and machine learning techniques. Different techniques are used to deal with continuous and discrete variables.

The main conclusions of the paper are that technological advance enabled to process huge amounts of data faster and in a more effective way and that Big Data analytics is becoming widely-used in many real-life applications nowadays and this is going to provoke rapid development of new statistical tools designed for dealing with bigger and more diverse sources of data. Also huge amount of data available now can lead to enhancement of previously known data processing techniques and to increasing their efficiency.