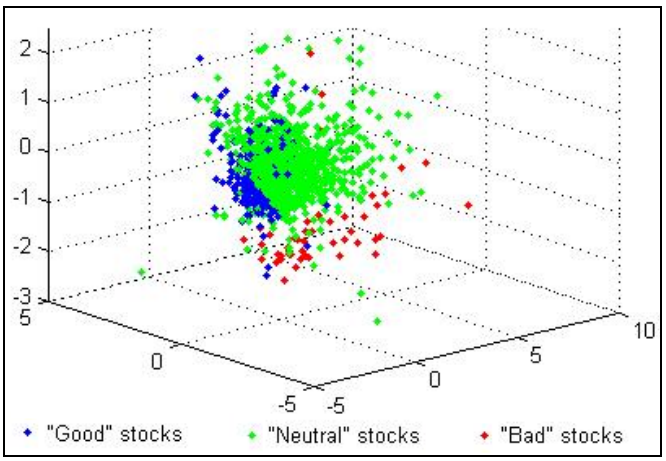

Fig. 9. Visualizing initial data set using MDS

V. CONCLUSIONS

Various data mining methods are used for examining large financial data sets to uncover hidden and useful information. This research focused on financial data visualization that is based on dimensionality reduction methods. We used data set that contained financial ratios of stocks traded on NASDAQ stock exchange. A brief overview of the most popular data dimensionality reduction and visualization methods was presented in this paper. We also showed how to adjust the algorithm of Random projection method for parallel computing. The MPI technology was applied in computer cluster to perform dimensionality reduction. The performance results revealed the advantages of parallel computing. Our goal was to visualize data and uncover hidden information. In order to do this we had to reduce the dimensions to 2 or 3. In the first step we executed Random projection algorithm in the cluster to reduce the dimensions from initial 54 to smaller amount. In the second step we applied Multidimensional scaling method to visualize the reduced data set.

One of the data set ratios was the recommendation of financial analysts. We raised hypothesis that companies having recommendation to buy them should be separated from the rest of companies. But the results showed that the combination of Random projection and Multidimensional scaling methods couldn't do this. This might have happened because in reality the potentiality of "good" and "bad" companies (as they are determined by analysts' recommendations) doesn't differ so much. However, in the second case of stock classification based on multicriteria indicator "good" and "bad" stocks were separated quite well. This leads to suggestion, that dimensionality reduction and visualization methods can effectively classify data and find the most promising stocks. But in order to explain the differences between classes we need to use several different ratios. It should be also noted, that MDS method alone was more accurate than combination of two methods.

REFERENCES

[1] R. Chowdhury, "Parallel Computing with OpenMP to solve matrix Multiplication," UCONN BIOGRID REU Summer 2010. Department of Computer Science & Engineering. University of Connecticut, Storrs, CT 06269.

[2] Dr. Domeniconi, "Comparison of Principal Component Analysis and Random Projection in Text Mining," April 29, 2004. INFS 795.

[3] I. K. Fodor, "A survey of dimension reduction techniques," Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, June 2002

[4] Intel Parallel studio. Access: https://software.intel.com/en-us/intel-parallel-studio-xe

[5] C. O.S. Sorzano, J. Vargas, A. Pascual Montano, "A survey of dimensionality reduction techniques," 2014. Access: arXiv:1403.2877.

[6] A. K. Menon. Random projections and applications to dimensionality reduction. School of Information Technologies, The University of Sydney, 2007.

[7] MPI technology. Access: http://www.mpi-forum.org/docs/mpi-3.1/mpi31-report.pdf

[8] H. Kim , P. Howland, H. Park. Dimension Reduction in Text Classification with Support Vector Machines. Journal of Machine Learning Research 6 (2005) 37–53

[9] S. Kudyba. Big Data, Mining, and Analytics–Components of Strategic Decision Making. March 12, 2014 by Auerbach Publications. Reference - 325 Pages - 89 B/W Illustrations. ISBN 9781466568709 - CAT# K16400

[10] Message passing interface. Access: https://computing.llnl.gov/tutorials/mpi/

[11] M. Mizuta, "Dimension Reduction Methods, Papers," Humboldt-Universität Berlin, Center for Applied Statistics and Economics (CASE), 15, 2004.

[12] R. S. Rosaria, I. Adae, A. Hart, M. Berthold, "Seven Techniques for Dimensionality Reduction," Knime, 2014.

[13] H. Jin, D. Jespersen etc. High performance computing using MPI and OpenMP on multi-core parallel systems. Parallel Computing 37 (2011) 562–575

[14] R. Rabenseifner, G. Hager, G. Jost, "Hybrid MPI/OpenMP Parallel Programming on Clusters of Multi-Core SMP Nodes," Conference: Proceedings of the 17th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, PDP 2009, Weimar, Germany, 18-20 Febuary 2009

[15] F. Cappello, D. Etiemble, "MPI versus MPI+OpenMP on the IBM SP for the NAS Benchmarks," Supercomputing, ACM/IEEE 2000 Conference. ISSN: 1063-9535.

[16] G. E. Dahl, J. W. Stokes, L. Deng, D. Yu, "Large-scale malware classification using random projections and neural network," Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference, pp. 3422-3426, 2013.

[17] E. Bingham, H. Mannila, "Random projection in dimensionality reduction: Applications to image and text data," Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 245-250, 2001.