

Data Farming: A Primer

Andrew Kusiak¹

Intelligent Systems Laboratory, Mechanical and Industrial Engineering, 2139 Seamans Center, The
University of Iowa, Iowa City, Iowa 52242-1527

Abstract—A typical data mining project uses data collected for various purposes, ranging from routinely gathered data, to process improvement projects, and to data required for archival purposes. In some cases, the set of considered features might be large (a wide data set) and sufficient for extraction of knowledge. In other cases the data set might be narrow and insufficient to extract meaningful knowledge or the data may not even exist. Mining wide data sets has received attention in the literature, and many models and algorithms for feature selection have been developed for wide data sets. Determining features for which data should be collected in the absence of an existing data set or when a data set is partially available has not been sufficiently addressed in the literature. Yet, this issue is of paramount importance as the interest in data mining is growing. The methods and process for the definition of the most appropriate features for data collection, data transformation, data quality assessment, and data analysis are referred to as data farming. This paper outlines the elements of a data farming discipline.

Keywords—Data farming, Data mining, Feature definition, Feature functions, New features

¹ Corresponding author's email: andrew-kusiak@uiowa.edu