Paper No: PU-SOE- CSE– 16

A Comparative Study of Statistical Analysis on Big Mart using Data Mining Techniques

T K Thivakaran 1 , Dr M Ramesh 2
1. Department of CSE, Presidency University, Bengaluru, India
2 Professor, Annamalai University, Tamil Nadu, India.

Abstract

In order to estimate sales revenue that is tangible and achievable, businesses involved in wholesales, manufacturing activities, marketing activities, retailing, logistics and supply chain activities need to use historical transaction data to forecast sales. In order to do this, there are several traditional data mining and statistical techniques that are used to identify trends, make predictive as well as descriptive analysis. The knowledge gained from such analysis is used in making business decisions. The data set in this study has been collected in the year 2013, and has 1559 products across 10 stores in different cities. First we conduct Exploratory Data Analysis to understand the nature of thAfter this, several traditional and novel data mining techniques have been applied on this data set, namely, linear regression, ridge regression, random forest regressor, decision tree regressor, XG Boost regressor and ARIMA. The cross-validation scores of all models are compared and inference as to which attributes and feature are given most weight during prediction of Item Outlet Sales attribute (target attribute) in the data set. Towards the end of the paper, the inferences and results are noted and discussed, hence completing the entire data analysis cycle.

Keywords:
Data Mining, Machine Learning, Gradient Descent, Gradient Boosting, Auto-regressive integrated moving average, Time-series data, Sales Prediction