Literature review:

Paper name: A support system for predicting ebay and prices

The predictors of the model: the number of pictures,the feedback rating,description of the item

Justifications for the predictors:

1. feedback rating:There are several studies indicating that there is a relationship between a seller’s reputation and the expected auction and price
2. Number of pictures:provides bidders with useful and objective information
3. Description of the item:making information publicly available raised the expected revenue for the seller

Machine learning model: boosting, text mining,decision tree

Data collection: web crawler or API or buying from third party

Data pre-processing:transform the description of an item into numerical values (using a dictionary, bags of words method)

Data set used: Canon digital camera, Nike Men shoes, H700 Motorola bluetooth headsets,30GB Apply ipod mp3 player

Data preparation:

1. check whether an item’s description influences the auction prices, we downloaded two homogeneous datasets and compare
2. Excluded outliers by filtering using box-plot of the item prices

Methodology and prediction model:

1. classification and regression tree ( decision trees are actually built in two phases. Growing phase and pruning phase)

The advantage of regression tree compared to linear regression:

1. . determine themselves which of the attributes are to be used for modelling the relationship with target variables.
2. Non-linear relationship with the target
3. Handle categorical data and missing value without transformation of data

Drawbacks: instability depend on the exact dataset for modelling therefore we use boosting

1. boosting: combine multiple models to improve performance

LSBoost Algorithm: mathematical foundation for general loss function

1. relative importance plots: visualize how important the various independent variables are related relative to one another in predicting the dependent variables

Experiments and results:

1. training sets and test sets (8/2)

2. evaluation of the model: mean absolute measure and mean relative measure

3.item description is the most significant predictor (eg:technical terms and condition of sales)