# PERFORMANCE COMPARISON OF PREDICTION MODELS FOR LED CHIP MANUFACTURING: A CASE STUDY

*Ya-Han Hu, Department of Information Management, National Chung Cheng University, 168 University Road, Min-Hsiung, Chia-Yi 62102, Taiwan, 886-5-2720411, yahan.hu@mis.ccu.edu.tw*
*Kuanchin Chen, Department of Business Information Systems, Western Michigan University, Kalamazoo, MI 49008, United States, 269-387-5795, kc.chen@wmich.edu*
*Chih-Chung Chou, Department of Information Management, National Chung Cheng University, 168 University Road, Min-Hsiung, Chia-Yi 62102, Taiwan, 886-5-2720411, choujj@mis.ccu.edu.tw*

## ABSTRACT

Due to its low energy consumption, long life, and small size, Light Emitting Diode (LED) technology has revolutionized the light industry. Today, over half of LED-related patents and technologies are owned by the large companies in United States, Japan, and Germany. In addition, the main devices for LED chip manufacturing are monopolized by United States and Germany. Under such a competitive environment, LED chip manufacturers elsewhere need to creatively find ways to increase yield and decrease the production cost to stay competitive.

The case company is one of the largest LED manufacturers in the world that produces customized LED chips for the global ODM clients. In LED chip manufacturing, the quality of LED chips is measured by its voltage and brightness. The aim of this study is to develop LED chip voltage prediction models to achieve the quality objectives of the company.

A successful prediction model is not only built from carefully selected learning algorithms, but also executed through a series of data processing tasks, including removal of duplicate records, and records with missing values, as well as data transformation and normalization. A number of supervised learning techniques were studied and compared, including linear regression (LR), model tree (MT), and support vector regression (SVR). To achieve higher prediction accuracy, we further consider classifier ensembles (i.e. bagging) for all single classifiers. In all experimental evaluations, the ten-fold cross validation technique was adopted to build the training and testing data sets. The outcomes of the above prediction models are then compared with a widely accepted set of metrics, such as mean absolute error (MAE), root mean squared error (RMSE), and correlation coefficient (CC).

The Correlation Coefficients of LR, MT, and SVM, are 0.3686, 0.7998, and 0.3515, respectively; the MAE of LR, MT, and SVM, are 0.0666, 0.0399, and 0.0662, respectively; the MAE of LR, MT, and SVM, are 0.0870, 0.0564, and 0.0876, respectively. These results show that regardless of the type of machine learning techniques, the prediction accuracy of LED chip voltage is promising. Of all the

algorithms in this study, MT had the best performance in prediction, while SVM had the worst performance. Results of using the bagging technique are: the CC of LR, MT, and SVM, are 0.3693, 0.8139, and 0.3527, respectively; the MAE of LR, MT, and SVM, are 0.0666, 0.0380, and 0.0662, respectively; the MAE of LR, MT, and SVM, are 0.0869, 0.0549, and 0.876, respectively. This shows that the model tree with bagging (Bagged MT) can improve prediction accuracy, but bagging had a limited effect on the performance of the other two prediction methods. Therefore, Bagged MT is suggested as the most effective prediction model, since it outperforms other algorithms in CC, MAE and RMSE. In addition, a number of critical factors affecting LED chip voltage are also identified. Our experimental evaluations show that machine learning techniques can help not only identify the critical factors that influence LED chip voltage, but accelerate the LED manufacturing yield.