

A Classification Approach with Different Feature Sets to Predict the Quality of Different Types of Wine using Machine Learning Techniques

Satyabrata Aich¹, Ahmed Abdulhakim Al-Absi², Kueh Lee Hui³, John Tark Lee³ and Mangal Sain⁴

¹Department of Computer Engineering, Inje University, South Korea

²Department of Computer Engineering, Kyungdong University- Global Campus, Gangwondo, South Korea

³Dept. of Electrical Engineering, Dong-A University, South Korea

⁴Department of Computer Engineering, Dongseo University, South Korea

satyabrataaich@gmail.com, absiahmed@kduniv.ac.kr, leehkueh@dau.ac.kr, jtlee@dau.ac.kr, mangalsain1@gmail.com

Abstract— In the past few years, with the availability of lot of wine brands it is difficult to identify the good quality wines. Good quality wine depends on the so many important factors such as chemical, scientific as well as technical factors. However in the previous study the researchers always focus on the subjective study to define the quality of wine. The result based on the subjective study takes much time as well as it is not effective compared to the objective study with the analytical methods. In the last few year's machine learning techniques caught lot of attention in every field. Most of the machines learning techniques are able to produce highly accurate result that compels most of the data scientist to implement it in case of predictive analytics. In the past few works related to wine data has been studied using different classifiers, however so far nobody has compared the performance metrics of the different classifiers with different feature sets to predict the quality of different type of wine by considering several factors. In this paper a new approach has been proposed by considering different feature selection algorithm such as Principal Component Analysis (PCA) as well as Recursive Feature Elimination approach (RFE) approach for feature selection and nonlinear decision tree based classifiers for analyzing the performance metrics. We found accuracies ranging from 94.51% to 97.79% with different feature sets using Random Forest classifier. This analysis will help the wine experts to know the important factors to consider while selecting the good quality wine.

Keyword— machine learning; feature selection; classifiers; performance metrics; wine quality



Satyabrata Aich is working as a researcher in the field of computer engineering. He has over four years of teaching, research and industry experience in India and abroad. He has published many research papers in journals and conferences in the realms of Supply Chain Management and data analytics. His research interests are natural language processing, Machine learning, supply chain management, data mining.



Ahmed Abdulhakim Al-Absi is an assistant professor in Department of Computer Engineering (Smart Computing) at Kyungdong University in South Korea. He earned a Ph.D. in computer science from Dongseo University in 2015. He received M.Sc. degree in information technology at University Utara Malaysia in 2011, and B.Sc. degree in computer applications at Bangalore University in 2008. His research interests include Big Data processing, Hadoop, Cloud computing, IoT, Distributed systems, Parallel computing, Bioinformatics, Security, and VANETS.



Kueh Lee Hui is working as an assistant professor at the department of Electrical Engineering, Dong-A University since 2012. She completed her PhD Degrees from Department of Electrical Engineering, Dong-A University, Korea. In 2009 she completed her BS degree in Electronic and Communication, Department of Electronic Engineering, University Malaysia of Sarawak, Malaysia. She also has done MS in 2007 from Malaysia. Her research interests are image processing, face recognition, digital image forensic, intelligent control and control application, power system.

**John Tark Lee**

1979 : BS degree in Electrical Engineering of Dong-A University.

1981 : MS degree in Electrical Engineering of Dong-A University.

1988 : PhD degrees in Electrical Engineering of Chung-Ang University.

1983 ~ 1985 : A researcher in LG ELelectronic Co. Ltd.

1985 ~ Present : Professor, Department of Electrical Engineering, Dong-A University

Research interests : fuzzy theory, neuron-science, GA, intelligent controller design including nonlinear stability analysis.



Mangal Sain received the M.Sc. degree in computer application from India in 2003 and the Ph.D. degree in computer science in 2011. Since 2012, he has been an Assistant Professor with the Department of Computer Information Engineering, Dongseo University, South Korea. His research interest includes wireless sensor network, cloud computing, Internet of Things, embedded systems, and middleware. He has authored over 50 international publications including journals and international conferences. He is a member of TIIIS and a TPC member of more than ten international conferences.