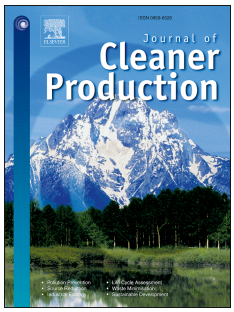
Accepted Manuscript



Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees

Muhammad Waseem Ahmad, Jonathan Reynolds, Yacine Rezgui

PII: S0959-6526(18)32555-1

DOI: 10.1016/j.jclepro.2018.08.207

Reference: JCLP 13991

To appear in: *Journal of Cleaner Production*

Received Date: 30 April 2018

Revised Date: 17 July 2018

Accepted Date: 19 August 2018

Please cite this article as: Ahmad MW, Reynolds J, Rezgui Y, Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees, *Journal of Cleaner Production* (2018), doi: 10.1016/j.jclepro.2018.08.207.

Total wordcount: 8367

Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees

Muhammad Waseem Ahmad[a,*], Jonathan Reynolds[a], Yacine Rezgui[a]

[a]*BRE Centre for Sustainable Engineering, School of Engineering, Cardiff University, Cardiff, CF24 3AA, United Kingdom*

Abstract

Predictive analytics play an important role in the management of decentralised energy systems. Prediction models of uncontrolled variables (e.g., renewable energy sources generation, building energy consumption) are required to optimally manage electrical and thermal grids, making informed decisions and for fault detection and diagnosis. The paper presents a comprehensive study to compare tree-based ensemble machine learning models (random forest – RF and extra trees – ET), decision trees (DT) and support vector regression (SVR) to predict the useful hourly energy from a solar thermal collector system. The developed models were compared based on their generalisation ability (stability), accuracy and computational cost. It was found that RF and ET have comparable predictive power and are equally applicable for predicting useful solar thermal energy (USTE), with root mean square error (RMSE) values of 6.86 and 7.12 on the testing dataset, respectively. Amongst the studied algorithms, DT is the most computationally efficient method as it requires significantly less training time. However, it is less accurate (RMSE = 8.76) than RF and ET. The training time of SVR was 1287.80 ms, which was approximately three times higher than the ET training time.

Keywords: Artificial intelligence; Extra trees; Random Forest; Decision trees; Ensemble algorithms; Solar thermal energy systems;

1. Introduction

The existing building sector, which is one of the most substantial consumers of energy, contributes towards 40% of world's total energy consumption and accounts for 30% of the total CO_2 emissions [1]. Currently, energy systems are predominantly based on fossils fuels. However, to reduce CO_2 emissions and tackle the challenge of mitigating climate change, such systems need to include a combination of fluctuating renewable energy resources (RES) such as wind and solar energy, along with residual resources (e.g., biomass) [2]. In recent years, more focus is being placed on increasing the energy efficiency, incorporating generation renewable energy sources and optimally managing the fluctuation of energy supply [3]. Energy generation through direct harnessing of solar radiation is one of the largest renewable energy technologies currently exploited worldwide. Solar energy currently constitutes a significant proportion of renewable energy generation in the EU. The majority of this energy generation is currently harnessed through solar photovoltaic systems for producing electricity, accounting for around 4.3% of total installed renewable energy in the EU in 2016 [4]. In contrast solar thermal energy only accounts for around 2% of installed renewable generation. To ensure a renewable energy future, it is vital that heating and cooling demands are

*Corresponding author

Email addresses: AhmadM3@Cardiff.ac.uk (Muhammad Waseem Ahmad), ReynoldsJ8@Cardiff.ac.uk (Jonathan Reynolds), RezguiY@Cardiff.ac.uk (Yacine Rezgui)