ROBERT GORDON UNIVERSITY ABERDEEN
SCHOOL OF COMPUTING SCIENCE AND DIGITAL MEDIA

INFORMATICS INSTITUTE OF TECHNOLOGY

CMM799 - MSc Project - BA

Student Name: Bhagya Antoinet Wickramasinghe	**Matriculation Number:** 2118340 /20210925
Supervisor: Mr. Dinesh Asanka	**Second Marker:**
Project Title: Forecasting of Medium-Term Energy Output of On-Grid Rooftop Photovoltaic Arrays – Case study for a Sri Lankan Solar Panel Installer	
MSc Business Analytics (CMM799) **Word Count:** 19,248	**Start Date: 2nd of October 2023**
	Submission Date: 5th of May 2023

CONSENT

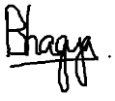
I agree ☒
I do not agree ☐

That the University shall be entitled to use any results, materials or other outcomes arising from my project work for the purposes of non-commercial teaching and research, including collaboration.

DECLARATION

I confirm:

- **That the work contained in this document has been composed solely by myself and that I have not made use of any unauthorised assistance.**
- **That the work has not been accepted in any previous application for a degree.**
- **All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.**

Student Signature: 	Date Signed: 29/04/2023

Abstract

The world is shifting towards the higher utilization of renewable energy sources in the road to greener energy, conserving an environmentally friendly atmosphere. The generation of sustainable energy via adopting solar photovoltaic is common worldwide. In the Sri Lankan context, a community-based project named "Soorya Bala Sangramaya" (Battle for Solar Energy) has been initiated by the ministry of Power to aid the community to cater their own power demand via rooftop solar panels.

The objectives of the research study were to identify the salient factors contributing to the energy generation of PV systems, to utilize a gamut of machine learning algorithms to build the predictive model and to identify the best machine learning algorithm to predict the energy generation based on accuracy and precision metrices. These objectives aids to achieve the aim of this study, which is to build a predictive model to determine the medium-term energy generated from on-grid rooftop PV systems.

The study has unveiled a new piece of knowledge on how the photovoltaic system dynamics and location specific data has contributed to the prediction of the system's power output. Further the findings are of paramount importance to the industry experts as well as the current and prospective solar panel users.

The data of all installer's solar panel sites was utilized and extracted from the source information systems. The necessary transformations and validations were applied and a detailed analysis was performed. The feature engineering, feature scaling (standardization), outlier-handling (standard deviation), multi-collinearity and feature selection was performed on data. The intended forecasting model based on fourteen supervised machine learning algorithms was built. The regression evaluation metrices explained variance score, maximum error, mean squared error, mean absolute error and coefficient of determination was utilized.

The KNN Regression algorithm in the factor analysis of all features after principal component analysis has outperformed all other built models. The best features comprises of PCA Component (inverter capacity category, solar panel capacity category, inverter capacity, solar panel capacity, inverter capacity per tilt and inverter capacity per month), tilt per solar panel capacity, longitude, azimuth per month, month per solar panel capacity, tilt per month, month, latitude, azimuth per solar panel capacity, inverter capacity per azimuth, inverter capacity per solar panel capacity, azimuth per tilt, model, azimuth category, installed direction, tilt category and brand. Moreover, a strong positive co-relation (0.92) was observed in the PCA component towards the solar panel energy output prediction.

As part of future work, building models utilizing a wider sample of on-grid roof top solar plants is imperative. Further it can be expanded to off-grid and hybrid solar plants along with roof mounted and ground mounted mounting methods. Moreover, it's vital to utilize other supervised machine learning algorithms to build the solution for the research problem under consideration.

Key Words: Photovoltaic, Forecast, Regression, Machine Learning Algorithm, Medium-Term