



الجمهورية الجزائرية الديمقراطية الشعبية
People's Democratic Republic of Algeria

وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research

المدرسة الوطنية العليا للتكنولوجيا والهندسة - عنابة -
National Higher School Of Technology And Engineering -ANNABA-

Department of Electronics, Electrical Engineering and Automation

End-of-Studies Project

In pursuit of the State Master's Degree

Field : Science and Technology

Program : Electrical Engineering

Specialty : Power Systems

Realized by

Mr. Smail MANSOURI

Ms. Sarra MADI

Advancing Photovoltaic System Performance Analysis through Data Visualization and Intelligent Algorithms

Supervised by

Dr. Azzeddine DEKHANE

Dr. Abdallah BOUZITOUNA

NHSTE

NHSTE

Thesis defended before a jury composed of :

Dr. Karima SMILI

Dr. Tarek MESIKH

Dr. Badri REKIK

President

Examiner

Examiner

NHSTE

NHSTE

NHSTE

Year 2024

Acknowledgement

"إِنَّمَا يَخْشَى اللَّهَ مِنْ عِبَادِهِ الْعُلَمَاءُ ﴿٢٨﴾" سورة فاطر

First and foremost, We would like to praise **Allah** the almighty, the most graceful, and the most merciful for his abundant blessings throughout our academic journey and the successful completion of this thesis. May Allah's blessings go to His final prophet **Muhammad** (peace be upon him), his family, and his companions.

We express our gratitude to **Dr. Azzeddine Dekhane** and **Dr. Abdallah Bouzitouna** for their essential counsel, inspiration, and assistance. Their deep expertise and perceptive criticism have greatly influenced the direction of this project. We sincerely appreciate all of their time and effort in going over our work and offering helpful critiques that made our study much better.

The combined mentorship and support of these individuals have not only contributed significantly to the completion of this project but have also enriched our academic and personal growth. We are truly fortunate to have had the opportunity to work with such distinguished professors and mentors.

Smail Mansouri & Sarra Madi

Dedication

“

I dedicate this work to

*My dear parents, whose love, sacrifice, and encouragement
have been the guiding light of my academic journey.*

*To my beloved brothers, Ayoub and Mohamed, your support
means the world to me.*

*In loving memory of my grandmother, whose presence I dearly
miss, yet whose teachings and love continue to resonate within
me every day. "اللَّهُمَّ ارْحَمْ جَدَّتِي وَأَغْفِرْ لَهَا"*

*To all my best friends, DRIHEM Youcef Rami, BELHACINI
Lina, MANAA Assil, SELMANE Anouar and all my beloved
companions, your support has been invaluable.*

”

Smail Mansouri

Dedication

“

I dedicate this work to

You, my steadfast pillars, my source of strength and support along the way, my parents. Every step of the way has been guided by your values and your inspiring example.

To my beloved sister Fairouz and my dear grandmother.

To my sister and best friend, the one who has supported me through all these years, in the good times as well as the most difficult, to the person dearest to my heart, my sister Hiba.

To all my best friends, BADJI Sabra, NAIEM Eya, BENGAOUER Chaima , BOUAZIZI Nour and all my beloved companions, your support has been invaluable.

”

Sarra Madi

Abstract

Accurate prediction of the maximum power point (P_{mpp}) of photovoltaic (PV) systems is crucial to optimize the energy yield and maximizing the efficiency of solar energy harvesting. This master thesis explores the potential of data-driven approaches for improving P_{mpp} prediction, utilizing regression techniques and feature importance analysis.

The study analyzed a dataset of irradiance, temperature, and P_{mpp} measurements, investigating the relationships between these variables and employing various regression models, including Ridge Regression, Lasso Regression, Decision Tree Regression, and Random Forest Regression. Performance comparisons revealed that tree-based models, notably Random Forest, outperformed linear models in capturing the complex interactions between input features and P_{mpp} .

Furthermore, feature importance analysis highlighted the significant influence of irradiance (G_{POA}) on P_{mpp} , particularly for tree-based models, underscoring the need for accurate irradiance data and modeling techniques that effectively capture non-linear relationships.

This master thesis concludes that data-driven approaches, specifically those employing tree-based models, hold significant potential for advancing P_{mpp} prediction and optimizing PV system performance. Future research should explore the integration of additional features, such as solar panel characteristics, atmospheric conditions, and system degradation factors, along with advanced machine learning techniques, to further enhance P_{mpp} prediction accuracy.

Keywords: Photovoltaic Systems, Maximum Power Point (P_{mpp}), Data Driven Approaches, Regression Techniques, Feature Importance Analysis, Irradiance (G_{POA}), Tree-Based Models, Nonlinear Relationships, Optimizing Energy Yield, Solar Energy Harvesting, Predictive Accuracy, Machine Learning.

ملخص

لتحسين عائد الطاقة وزيادة كفاءة جمع الطاقة الشمسية، يُعتبر التنبؤ الدقيق بنقطة القدرة القصوى (P_{mpp}) ضرورياً لأنظمة الطاقة الشمسية الفوتوفولتائية. تتناول هذه الدراسة استخدام البيانات لتحسين توقعات P_{mpp} باستخدام تقنيات الانحدار وتحليل أهمية السمات.

قامت الدراسة بتحليل مجموعة بيانات تشمل قياسات للإشعاع ودرجة الحرارة و P_{mpp} ، حيث استكشفت العلاقات بين هذه المتغيرات واستخدمت نماذج انحدار مختلفة مثل انحدار الريدج وانحدار اللاسو وشجرة القرارات والغابات العشوائية. أظهرت المقارنات بين الأداء أن نماذج الغابات العشوائية تتفوق على النماذج الخطية في التعامل مع التفاعلات المعقدة بين المتغيرات الداخلية و P_{mpp} .

بالإضافة إلى ذلك، كشف تحليل أهمية السمات عن تأثير كبير للإشعاع (G_{POA}) على P_{mpp} ، خاصة عند استخدام نماذج الغابات العشوائية، مما يبرز أهمية جمع البيانات الدقيقة للإشعاع وتوظيف تقنيات النمذجة التي تلتقط بفعالية العلاقات غير الخطية.

يستنتج البحث أن استخدام البيانات، وخاصة النماذج المبنية على الأشجار، يحمل إمكانات كبيرة لتقدم في توقعات P_{mpp} وتحسين أداء أنظمة الطاقة الشمسية الفوتوفولتائية. ينبغي للأبحاث المستقبلية استكشاف دمج سمات إضافية مثل خصائص الألواح الشمسية وظروف الغلاف الجوي وعوامل التدهور، بالإضافة إلى تطبيق تقنيات التعلم الآلي المتقدمة لزيادة دقة توقعات P_{mpp} .

الكلمات الرئيسية: أنظمة الطاقة الشمسية الفوتوفولتائية، نقطة القدرة القصوى (P_{mpp})، النهج المبنى على البيانات، تقنيات الانحدار، تحليل أهمية السمات، الإشعاع، (G_{POA}) النماذج المبنية على الأشجار، العلاقات غير الخطية، تحسين عائد الطاقة، جمع الطاقة الشمسية، الدقة التنبؤية، التعلم الآلي.

Résumé

La prédiction précise du point de puissance maximale (P_{mpp}) des systèmes photovoltaïques (PV) est cruciale pour optimiser le rendement énergétique et maximiser l'efficacité de la collecte d'énergie solaire. Cette thèse explore le potentiel des approches basées sur les données pour améliorer la prédiction de P_{mpp} , en utilisant des techniques de régression et une analyse de l'importance des caractéristiques.

Ce mémoire de master a analysé un ensemble de données comprenant des mesures d'irradiance, de température et de P_{mpp} , en étudiant les relations entre ces variables et en utilisant divers modèles de régression, notamment la régression Ridge, la régression Lasso, la régression par arbres de décision et la régression par forêts aléatoires. Les comparaisons de performances ont révélé que les modèles basés sur les arbres, notamment les forêts aléatoires, surpassaient les modèles linéaires en capturant les interactions complexes entre les caractéristiques d'entrée et P_{mpp} .

De plus, l'analyse de l'importance des caractéristiques a souligné l'influence significative de l'irradiance (G_{POA}) sur P_{mpp} , en particulier pour les modèles basés sur les arbres, soulignant ainsi la nécessité de données d'irradiance précises et de techniques de modélisation efficaces pour capturer les relations non linéaires.

Ce mémoire de master conclut que les approches basées sur les données, en particulier celles utilisant des modèles basés sur les arbres, ont un potentiel significatif pour faire avancer la prédiction de P_{mpp} et optimiser les performances des systèmes PV. Les recherches futures devraient explorer l'intégration de caractéristiques supplémentaires telles que les caractéristiques des panneaux solaires, les conditions atmosphériques et les facteurs de dégradation du système, ainsi que des techniques avancées d'apprentissage machine, pour améliorer davantage la précision de la prédiction de P_{mpp} .

Mots-clés : Systèmes photovoltaïques, Point de puissance maximale (P_{mpp}), Approches basées sur les données, Techniques de régression, Analyse de l'importance des caractéristiques, Irradiance (G_{POA}), Modèles basés sur les arbres, Relations non linéaires, Optimisation du rendement énergétique, Collecte d'énergie solaire, Précision prédictive, Apprentissage machine.

List of Abbreviations and Acronyms

PV Photovoltaic

GPOA Global Plane of Array Irradiance

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

MSE Mean Squared Error

R² Coefficient of Determination

List of Tables

Chapter I

I.1 Data Overview 4

Chapter IV

IV.1 Comparison of Regression Model Performance 23

List of Figures

Chapter I

I.1	photovoltaic system	2
I.2	I-V/P-V curves	3

Chapter III

III.1	Decision Tree Learning Curve	13
III.2	Lasso Regression Learning Curve	14
III.3	Ridge Regression Learning Curve	14
III.4	Random Forest Regression Learning Curve	14
III.5	Ridge Regression	14
III.6	Lasso Regression	14
III.7	Random Forest Regression	15
III.8	Decision Tree Regression	15
III.9	Actual vs Predicted (Ridge Regression)	16
III.10	Actual vs Predicted (Decision Tree Regression)	16
III.11	Actual vs Predicted (Lasso Regression)	16
III.12	Actual vs Predicted (Random Forest Regression)	17

Chapter IV

IV.1	Correlation Heatmap	18
IV.2	Correlation Coefficients	19
IV.3	Scatter Plots	19
IV.4	Scatter Plot of T_{MOD} and GPOA vs P_{mpp}	19
IV.5	Importance of features for Ridge regression	21
IV.6	Importance of features for Lasso regression	21
IV.7	Importance of features for Decision Tree regression	21
IV.8	Importance of features for Random Forest regression	21
IV.9	MSE Result	22

IV.10 Coefficient of Determination	22
IV.11 Comparison plot	22

Contents

Acknowledgement	I
Dedication	II
Abstract	IV
ملخص	V
Résumé	VI
List of Abbreviations and Acronyms	VII
List of Tables	VIII
List of Figures	X
General Introduction	1
I Chapter I: Context and Problem Statement	2
I.1 Introduction	2
I.2 Data Overview and Significance of P_{mpp} Prediction	3
I.2.1 Description of the Data	3
I.2.1.1 Interest and Exploitation of the Data	4
I.2.2 Importance of Predicting P_{mpp}	4
II Chapter II: Analysis and Implementation of Mathematical Methods	5
II.1 Mathematical Methods	5
II.1.1 Ridge Regression Method	5
II.1.1.1 Understanding the L2 Penalty	5
II.1.1.2 Advantages of Ridge Regression	5
II.1.2 Decision Tree Regression Method	6
II.1.2.1 Understanding Decision Trees for Regression	6
II.1.2.2 Advantages of Decision Tree Regression	6
II.1.3 Random Forest Regression	7
II.1.3.1 Understanding Random Forests for Regression	7
	XI

II.1.4	Lasso Regression	8
II.1.4.1	Understanding the L1 Penalty	8
II.1.4.2	Advantages of Lasso Regression	8
III	Chapter III: Regression Algorithms for P_{mpp} Estimation: A Comprehensive Evaluation	9
III.1	Comparative Analysis of Four Regression Methods	9
III.1.1	Ridge Regression Method	9
III.1.1.1	Ridge Regression Pseudocode	9
III.1.2	Decision Tree Regression Method	10
III.1.2.1	Decision Tree Regression Pseudocode	10
III.1.3	Random Forest Regression Method	11
III.1.3.1	Random Forest Regression Pseudocode	11
III.1.4	Lasso Regression Method	11
III.1.4.1	Lasso Regression Pseudocode	12
III.2	Advantages of Mathematical Methods	12
III.3	Limitations and Challenges of Mathematical Methods	13
III.4	Learning Curves Analysis	13
III.4.1	Decision Tree Model	13
III.4.2	Lasso Regression Model	14
III.4.3	Ridge Regression Model	14
III.4.4	Random Forest Regression	14
III.5	Evaluation of Predictive Accuracy	14
III.6	Actual vs Predicted Outputs	15
IV	Chapter IV: Results & Discussion	18
IV.1	Introduction	18
IV.2	Linear Relationship Analysis of P_{mpp} , T_{MOD} , and GPOA	18
IV.3	Feature Importance Analysis	20
IV.4	Results analysis	22
IV.5	Conclusion	23
	General Conclusion	25
	Bibliography	26