An IoT-based Hedge System for Solar Power Generation

Jia-Hao Syu, Mu-En Wu, Gautam Srivastava, Chi-Fang Chao, and Jerry Chun-Wei Lin*

***Abstract*—Environmental protection is an important issue in recent decades, and renewable energy is an ideal solution for eco-friendly power generation. Solar-power generation is a popular renewable energy with low cost and small environmental footprint, which leads to exponential growth and high industrial investment. A mature solar business model has been established, but some uncertainties hinder the development, especially when focusing on the lack of solar-radiation. To address these issues, in this paper we propose a hedging system to hedge the low-radiation risk for solar-investors through the designed IoT-based data, edge-based models for predicting solar-radiation as well as hedging options. Our experimental results show that the edge-based predictive models can obtain an R-squared value of 0.841 and a correlation coefficient of 0.917. For binary options designed in the hedging system, the broker can obtain stable payoffs with the highest Sharpe ratio of 3.354, and the investors can obtain large payoffs during low-radiation. Our simulation results show the effectiveness of the proposed hedging system for investors (buyer-side), simultaneously, present the motivation of the broker (seller-side) to join the designed hedging system utilized in solar-power generation.**

***Index Terms*—Solar-power generation, hedging, IoT-based model, edge computing, machine learning**

I. Introduction

ENVIRONMENTAL protection is an important and thought-provoking issue for researchers and industry alike in recent decades. Due to global warming and climate change, the means to generate electricity has become a major topic of research and development in recent years. Currently, most electricity comes from thermal power which produces large amounts of carbon dioxide (CO_2, greenhouse gases) and other harmful gases. These pollutants are the prime culprits for global warming [1]. Furthermore, pollutants from thermal energy have residual effects on the body including harming the lungs [2]. Renewable green energy is an ideal solution for environmentally friendly (eco-friendly) power generation [3], which can include wind-power, hydropower, and solar-power [4], [5]. Wind, as well as hydropower, utilize kinetic energy and water respectively to drive power generators. However, they are location-dependent, require high cost, and need ample space to operate. Conversely, solar-power generation has low cost and small footprint characteristics, and only a few square feet of power generation panels are needed to collect solar energy and generate a usable amount of electricity. These advantages make solar-power generation a viable option for green energy which has seen an increase in investment interest for companies and investors.

Simultaneously, many governments encourage investment and the construction of solar power generation, as well as provide substantial subsidies and guarantees to maintain a stable purchase price for green energy [6]. Lei *et al.* showed that 40% of global electricity growth comes from renewable energy sources, mainly from solar-power (40.7%), and wind-power (58.2%) [6]. This shows that solar-power generation is an emerging topic. With the unremitting efforts of both government and entrepreneurs, numerous people may invest in solar-power generation, and a mature solar business model has been established, as shown in Fig. 1. Companies that run solar-power plants are known to construct power generation infrastructure, which divides the power plant into shares to investors. Then, the generated electricity is sold to the government at a guaranteed stable price. For solar-power investors, they provide funds (investments) to build power plants and obtain shares. A share represents several units of power panels, and the electricity (profit) generated by these panels are distributed to investors with shares.

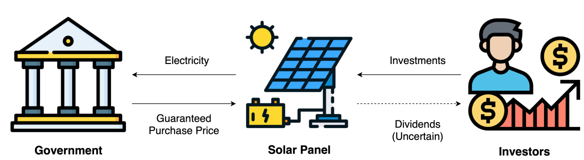


Fig. 1: The existing business model of solar-power generation

With a tremendous amount of information in various regions and interaction between parties, techniques for the Internet of Things (IoT) [7], data mining, and edge computing are suitable technologies to utilize in these circumstances. One of the main concepts behind IoT is to exchange virtual information and interact with physical objects and smart devices [8], [9] through sensing, control, mining [10], and the concern of security [11], [12]. Data mining and machine learning are usually adopted in IoT systems to analyze information. Data mining aims to discover structures, patterns, as well as information within

Jia-Hao Syu is with the Department of Computer Science and Information Engineering, National Taiwan University, Taiwan. Email: k1143001@gmail.com

Mu-En Wu is with the Department of Information and Finance Management, National Taipei University of Technology, Taiwan. Email: mnwu@ntut.edu.tw

Gautam Srivastava is with the Department of Mathematics and Computer Science, Brandon University, Brandon, Canada as well as with the Research Center for Interneural Computing, China Medical University, Taichung 40402, Taiwan. Email: srivastavag@brandonu.ca

Chi-Fang Chao is with the Department of Information and Finance Management, National Taipei University of Technology, Taiwan. Email: s9860320@gmail.com

Jerry Chun-Wei Lin is with Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway. Email: jerrylin@ieee.org (*Corresponding author)

datasets [13], [14] through data- and demand-driven models [15]. In contrast, machine learning aims to build algorithms for classification and regression through sampled data and past experiences to predict and assist decision-making [16], [17]. Since IoT-based data is usually located in various regions, edge computing can be implemented to distribute computational efforts to the edges of IoT systems [18], while attempting to address known concerns at the edge of networks such as response time, data privacy [19], throughput, and energy efficiency [20], [21]. Through ubiquitous embedded and edge systems and data mining techniques, IoT systems can make at to the ease with which we live today [22], [19] by establishing smart cities [23], smart healthcare [24], smart agriculture [25], and many other technologically advanced infrastructures.

Although the above business model can give some level of guarantee for earned profits from the government, there are still some uncertainties including natural disasters and lack of solar-radiation. Damages caused by natural disasters can be compensated through insurance (property insurance has been well-developed [26]). The only uncertainty and uncontrollable factor in the solar business model is solar-radiation, which completely dominates the electricity generated by solar panels. Less solar-radiation will produce less electricity and profits, which may not be able to cover the depreciation and result in investment losses. To address these issues, we propose a novel hedging system utilized in the solar-power business, which adopts the edge-based predictive models with IoT data for solar radiation, and the hedging binary option. IoT-based data contains information from solar panels and weather sensors. The designed two edge-based predictive models are constructed with four well-developed machine learning techniques to achieve distributed computing with low hesitation and less computation. The binary options act as an intermediary for the hedging service.

In the developed hedging system, binary options act as an intermediary to hedge against low solar-radiation risk. Investors can purchase options in our hedging system (for example, bet that the radiation is less than 20 J/m^2, joule per square meter) to hedge low-radiation risk. If solar-radiation is less than 20 J/m^2, the investors may have losses in a solar-power investment but can earn a payoff from the options. If the radiation is greater than 20 J/m^2, investors only need to spend a small number of finances for hedging and obtain more profits from their solar-power generation. The broker (seller-side of the option) must accurately determine the odds of each option, which is determined by predictive solar-radiation and the probability of binary outcomes. Therefore, a precise prediction is required in the proposed system to accurately predict solar-radiation, which is called the precise predictive model (PPM). To further process real-time information and solve low-latency issues on our edge-based model, a light predictive model (LPM) is developed that to speed up the runtime performance with fewer features on the edge-computing model.

Our in-depth experimental results show that the prediction algorithm of random forest regression has the best performance, which obtains an R-squared of 0.841 (0.828) and correlation coefficient of 0.917 (0.910) within PPM (LPM). Robustness results on different datasets and comparing with state-of-art works also present the outstanding performance of the predictive models (PPM and LPM). Besides, the experimental results of hedging options show stable payoffs (with the highest Sharpe ratio of 3.354) for the broker and effective hedging services for investors. These results demonstrate the effectiveness of the proposed hedging system. Thus, the major contributions of this paper are then summarized as:

1) Developed an IoT-based data-driven system utilized in solar-power generation and prediction.
2) Adopted four machine learning algorithms into the two edge-computing models (PPM and LPM) to respectively predict solar-radiation accurately, and with low-latency and less computation.
3) Showed that the designed PPM and LPM obtained better performance in terms of mean square error and R-squared compare to the existing models.
4) Proposed hedging system provides hedging services for solar-power investors and generates fix-income for the broker in the simulation.

We organize this paper as follows. Section II reviews the literature on solar-power generation prediction, machine learning algorithms, and binary option. Section III presents the design of IoT-based data, predictive models for solar-radiation, and designed hedging option. Section IV first states the use of datasets, demonstrates the prediction results with robustness test, and evaluates the effectiveness of hedging options. Section V discusses the findings and summarizes the results of the developed models.

II. Literature Review

In this section, the background of solar-power generation and prediction are first introduced. Then, we survey the literature and machine learning algorithms used in this paper. Furthermore, the binary option is stated and discussed.

A. Solar-Power Generation and Prediction

Solar-power generation converts solar-radiation into electricity through the photovoltaic effect [27]. A single photovoltaic cell generates only a few watts of energy. By connecting an array of photovoltaic cells, the photovoltaic system (solar panel) can generate about 150 to 180 Watts per square meter [28]. The commonly photovoltaic system is the flat solar panel deployed in our daily lives (rooftop or bus station). Large solar-power plants may be implemented by solar trackers and concentrated solar power, which can rotate panels or concentrate sunlight to improve efficiency [29].

Partain *et al.* [30] illustrated that the cumulative capacity of solar power could be doubled every two years. Swanson's law stated that the price per watt of solar photovoltaic modules drops by half for every 10 times the capacity increases [31]. These pieces of evidence show that solar-power generation is the fast-growing renewable energy and is cost-competitive to the other thermal powers (e.g., coal, crude oil, and natural gas).

Zeng and Qiao [32] adopted a least-square support vector machine model (SVM) for predicting solar-power, which utilized features of historical atmospheric transmissivity and

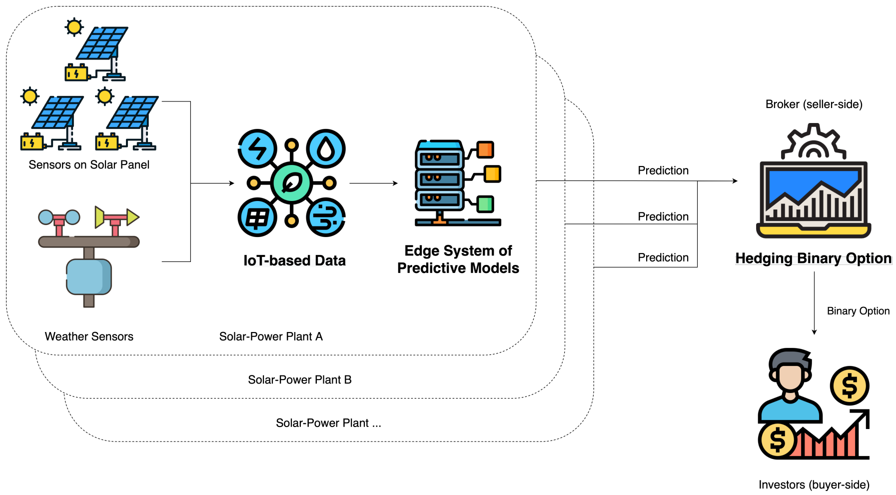


Fig. 2: The proposed hedging system utilized in solar-power generation

meteorological variables (e.g., sky cover, relative humidity, and wind speed). Compared with the models based on auto-regressive and neural networks, the SVM-based model achieved better results. Jang *et al.* [33] designed a model to predict the solar-power by atmospheric motion vectors from satellite images, which can be used to determine the motion vectors of clouds and affected the solar-radiation. They also adopted SVM as the prediction model and defeated the models with non-linear autoregressive and artificial neural networks. Long *et al.* [34] developed daily solar-power prediction models using data-driven approaches, and the artificial neural network-based model had the highest accuracy when predicting present, and multivariate linear regression-based and k-nearest neighbor-based models outperformed other models when predicting multi-steps.

TABLE I: Description and characteristics of the observations

Observation	Abbreviation	Description	Unit
Observation month	*Month*	the month of observation	Dummy, 1,...,12
Observation time	*Hour*	the time (which hour) of observation	Dummy, 1,...,24
Solar-radiation	*SRad*	the radiant energy of sunshine	MJ/m^2
Sun duration	*SDur*	the length of sunshine	hour
Temperature	*Temp*	average temperature during the observation	°C
Relative humidity	*RH*	relative humidity during the observation	%
Wind speed	*WS*	average speed of wind during the observation	m/s
Cloud amount	*CA*	the amount (region) of cloud cover the sky	0,...,10

B. Machine Learning

Machine learning is a group of powerful computational methods that can make accurate predictions through experiments [16], [35], [36]. To meet the requirements of the proposed edge system, we survey several efficient and light-computing algorithms, including multiple linear regression, support vector machine, and random forest regression.

1) Multiple Linear Regression: Multiple Linear Regression (MLR) is a linear method to estimate the relationship between a dependent variable and multiple dependent variables, which is also known as the level of correlation [37]. MLR is a linear model, and generally fitted by least-squares approach [38]. MLR is the most commonly used regression model for data analysis with the characteristics of easy-to-use and interpretable. Each independent variable is related to the dependent variable through its regression coefficient, which makes users explore relationships intuitively.

2) Support Vector Regression: Support Vector Machine (SVM) is a supervised learning algorithm, proposed by Cortes and Vapnik in late 1995, which is originally developed for two-type classification problems [39]. SVM aims to find a hyperplane with the largest margin to separate different categories of data through the features. The data on the margins provide the most information, which is called the support vector. Besides, techniques of nonlinear mapping can enhance the classification ability by separating data in high-dimensional spaces, and the mapping can be achieved through different kernel functions, which control the mapping and boundary shape. SVM is also extended to regression problems, especially nonlinear regression, called Support Vector Regression (SVR) [40]. Note that the computational complexity of SVR is approximately $\mathcal{O}(N^2)$ due to the computation of quadratic programming, where N is the number of training data.

3) Random Forest Regression: Random forest regression is a supervised learning algorithm developed by Leo Breiman in 2001 [41], which can be regarded as the expansion and aggregation of the decision tree by randomly sampling data and features with replacement. Each decision tree is a weak classifier, and multiple subsets of decision trees constitute a strong classifier, called the random forest. Through bagging and bootstrap techniques, random forests can achieve accurate and stable performance than the generic decision trees [42]. The advantage of the random forest is that it has a stronger generalization and classification ability to handle a large number of input features. Besides, feature selection is not necessary for random forests while dealing with high-dimensional data, which can seriously reduce the computational cost [43].

4) Multilayer Perceptron: Multilayer perceptron (MLP) is the basic method of feedforward artificial neural networks, which attempts to simulate the function of the human brain and the interaction between neurons [44]. MLP with multiple hidden layers is also known as deep learning [45]. MLP can be divided into three kinds of layers, including input, hidden, and output layers. The input and output layer controls the shape and type of input and output. As for the hidden layer, it is composed of perceptrons to perform calculations (through input values, trainable weights, and activation functions) and transfer the values to subsequent layers. With a high degree of freedom (layers, neurons, and activation functions), MLP is a powerful, flexible, and widely-used model for solving various problems [44].

C. Binary Options

Binary options are a common financial product, also known as a binary bet or fixed odds bet [46], [47]. This option regulates the underlying target, expiration time, binary condition, and the odds. At expiration time, if the status of the underlying target meets the binary condition of the binary option, the option holder can receive the face value (the amount he/she bet) multiplied by the odds. For example, binary option bet on whether team A defeats team B (binary condition) in the basketball game on September 15 (underlying target and expiration time), and the odds of the option is 2.85. Suppose a fan of team A buys this binary option for 100 and if team A defeats team B (the winning condition), they will be awarded 285 dollars (face value of 100 times odds of 2.85); otherwise, the fan will lose all of the 100 dollars.

Several positive characteristics of binary options make it a widely and heavily used financial product and sports bets, including flexibility, limited risk, and easy access [47]. Various underlying targets and binary conditions lead to the flexibility of binary options. Simple and controllable mechanisms (face value and odds) make it easy access and hedge risk and become the main reason for investor interest [46].

III. Proposed IoT-based Data-Driven Hedge System

The flowchart of the proposed hedging system is shown clearly in Fig. 2. Solar-power plants build solar panels and weather sensors to provide IoT-based information. The IoT-based data is then transmitted to the predictive edge-computing models to real-time predict the solar-radiation with distributed computation and achieving low-hesitation and less computation cost. Precise and light predictive models (PPM and LPM) are designed using machine learning algorithms to make accurate and efficient predictions. The predicted solar radiation is then transmitted to hedging binary options to calculate the odds of each option. Then, investors can purchase the hedging service of binary options according to the given odds to hedge against low solar-radiation risks. Three major modules in the proposed system include IoT-based data, edge-based predictive models, and hedging binary options, and are introduced in the following subsection.

A. IoT-based Data Collection

The IoT-based data collected from solar-power plants play an important role in the proposed hedging system. In the designed IoT-based data collection, each solar-power plant-equipped sensor on a given solar panel is used to collect information on generated electricity and received solar-radiation. Furthermore, meteorological information collected by self-built weather sensors in solar power plants (or external weather stations) is also collected. This information is physically collected via sensor networks, which will be aggregated together as the IoT-based data and transferred to the edge-based predictive models.

TABLE II: Features usage for predicting solar-radiation at h o'clock on date d ($SRad_{d,h}$)

Date (d)	$d-5$, $d-4$, $d-3$, $d-2$	$d-1$	d
SRad	$SRad_{d-5,h}$, ..., $SRad_{d-2,h}$	$SRad_{d-1,h}$	
SDur		$SDur_{d-1,h}$	
Temp		$Temp_{d-1,h}$	
RH		$RH_{d-1,h}$	
WS		$WS_{d-1,h}$	
CA		$CA_{d-1,h}$	
Month			$Month_{d,h}$
Hour			$Hour_{d,h}$

A simple example of the formed IoT-based data is shown in Table I. In each solar-power plant, received solar-radiation (*SRad*) is collected from sensors on solar panels, which is the prediction and hedging target of the proposed system. Then, collect weather information as prediction features, including temperature (*Temp*), humidity (*RH*), and wind as prediction features (*CA*). In summary, the IoT-based data is formed with the features of meteorological information and labels of solar-radiation.

B. Edge-based Predictive Models

The proposed hedging system would predict the amount of solar-radiation of each period (hour or day). Even if the final implementation of the system is on the binary options; however, there would be multiple thresholds to define the binary condition (low-radiation) of the options. If the classification models are applied, the system will have to fix the threshold of the option (less flexibility for investors) or have to train and predict on multiple models for multiple thresholds (more working-load and computation). Therefore, in this paper, we utilize the regression model to predict a continuous value of solar radiation, and simple transformations are used to obtain the probabilities under different thresholds.

The predictive models on edge systems aim to predict solar-radiation through machine learning algorithms with IoT-based data. The physical location of the predicting edge system is designed in the solar-power plant itself. Through distributed edge computing, lower response time and a higher privacy level can be achieved. Moreover, only an encrypted predicted value needs to be transmitted, which greatly reduces both transmission size and time. Using the received IoT information, the pre-trained machine learning algorithms embedded in the edge system can predict the solar-radiation in real-time. Then, the predicted solar-radiation will be used to calculate the odds of the binary options in the further modules of hedging binary options.

To obtain good prediction results, four classic machine learning algorithms are adopted which include multiple linear regression (MLR), random forest regression (RFR), support vector regression (SVR), and multilayer perceptron (MLP). Since these models are designed to predict the hourly or daily solar-radiation, several hourly observations are adopted as the prediction features, including observation month and time, sun duration, temperature, relative humidity, wind speed (reference from [32], [33]). The description and characteristics of the observations are shown in Table I.

To precisely predict solar-radiation, the precise predictive models called PPM is proposed, which utilizes 12 features on MLR, RFR, SVR, and MLP algorithms. 12 features include *Month* and *Hour* of the prediction time, and *SDur*, *Temp*, *RH*, *WS*, *CA* in the previous day, and *SRad* in past five days. For example, to predict the solar-radiation at h o'clock on date d ($SRad_{d,h}$), we utilize $SRad_{d-1,h}$, $SRad_{d-2,h}$, $SRad_{d-3,h}$, $SRad_{d-4,h}$, $SRad_{d-5,h}$, $Month_{d,h}$, $Hour_{d,h}$, $SDur_{d-1,h}$, $Temp_{d-1,h}$, $RH_{d-1,h}$, $WS_{d-1,h}$, and $CA_{d-1,h}$ as 12 features in PPM and shown in Table II. Note that the blank cells mean that we do not use the corresponding feature (at that time) in the designed PPM model. To satisfy low hesitation and less computation for the edge computing environment overall, the light prediction model (LPM) is proposed, which utilizes only 5 features on MLR, RFR, SVR, and MLP algorithms. 5 features are the historical solar-radiation in the past five days. Briefly, to predict the solar-radiation at h o'clock on date d ($SRad_{d,h}$); we utilize ($SRad_{d-1,h}$, $SRad_{d-2,h}$, $SRad_{d-3,h}$, $SRad_{d-4,h}$, $SRad_{d-5,h}$) as 5 features for LPM and shown in the first row (*SRad*) of Table II.

C. Designed Hedging Binary Options

For the designed hedging services of binary options, two main specifications are stipulated, respectively named the winning conditions and the odds (*ODDS*). The winning conditions determine the underlying target, expiration time, and binary conditions. In the design binary option, the underlying target is the cumulative solar-radiation of the station (within one hour or one day), the expiration time is the end of the hour or the day, and the binary condition is whether the cumulative solar-radiation is less than the threshold T. If the solar-radiation is less than T (the winning condition), the investor can get a payoff of the purchase value multiplied by the *ODDS*. For example, an investor purchases $1,000$ dollars of option that bets the cumulative solar-radiation on July 1, 2020 with a threshold of $10\ \frac{MJ}{\mathrm{m}^2}$ and an odds of 2 ($T = 10\ \frac{MJ}{\mathrm{m}^2}$, *ODDS* = 2). If the cumulative solar-radiation on July 1, 2020 is less than $10\ \frac{MJ}{\mathrm{m}^2}$, the investor can get $1,000 \times 2 = 2,000$ dollars, otherwise, investors lose 1,000 dollars. In the design hedging option, the broker provides hedging options for daily cumulative solar-radiation with $T = 10, 15, 20\ \frac{MJ}{\mathrm{m}^2}$. The investors can find the *ODDS* of hedging options with different T, and determine the type (T) and amount of options they should buy (hedge). For a simple implementation, we stipulate that the only broker can sell binary options and only investors can buy options.

Furthermore, the $OODS$ of each option is provided by the broker, who uses the proposed predictive models to predict the solar-radiation (*Pre*) and converts it into the fair odds. We assume that solar-radiation comes from a normal distribution with a mean of *Pre* and a standard deviation of σ, where σ is the historical standard deviation of historical solar-radiation. With the predicted normal distribution, we can calculate the probability of solar-radiation bellows T (winning condition), which is also the win rate (*WR*) for option holders (investors). Note that there may be negative values in the normal distribution, and we will ignore the probability in the negative region, and normalize probability in the positive region to 1. Let $F(X)$ be the cumulative distribution function of a normal distribution with a mean of *Pre* and standard deviation of σ, where X is the solar-radiation. The win rate *WR* is calculated in Equation 1. The probability of value less than T and larger than 0 is $F(\frac{T-Pre}{\sigma}) - F(\frac{0-Pre}{\sigma})$, and the probability of value larger than T is $1-F(\frac{T-Pre}{\sigma})$. Then, we normalize the probability to one, and calculate the win rate as $\frac{F(\frac{T-Pre}{\sigma})-F(\frac{0-Pre}{\sigma})}{1-F(\frac{0-Pre}{\sigma})}$ Furthermore, the fair odds and *ODDS* can be calculated by the win rate as given in Equation 2.

$$\begin{aligned} WR &= \frac{F(\frac{T-Pre}{\sigma}) - F(\frac{0-Pre}{\sigma})}{(1-F(\frac{T-Pre}{\sigma})) + (F(\frac{T-Pre}{\sigma}) - F(\frac{0-Pre}{\sigma}))} \\ &= \frac{F(\frac{T-Pre}{\sigma}) - F(\frac{0-Pre}{\sigma})}{1-F(\frac{0-Pre}{\sigma})} \end{aligned} \tag{1}$$

$$\begin{aligned} &WR \cdot \text{fair odds} = (1 - WR) \cdot 1 \\ &WR \cdot (\text{fair odds} + 1) = 1 \\ &\Rightarrow \text{fair odds} = \frac{1}{WR} - 1 \\ &\Rightarrow ODDS = \text{fair odds} - \text{commission} \\ &\qquad\qquad = \frac{1}{WR} - 1 - \text{commission} \end{aligned} \tag{2}$$

Fairness means that the expected payoffs of winning and losing conditions should be equal. When an investor buys a dollar of option, the expected losing payoff is the probability of loss multiplied by one dollar, $(1 - WR) \cdot 1$. The expected winning payoff is the win rate multiplied by the fair odds $WR \cdot$fair odds. Therefore, $(1-WR) \cdot 1 = WR \cdot$fair odds, and we can get fair odds $= \frac{1}{WR} - 1$. To increase the willingness of the broker, conventionally, commissions are charged to subsidize the broker's fees and as a fixed income. Usually, commissions are directly deducted from the odds; therefore, the final odds (*ODDS*) is fair odds − commission.

IV. Experimental Results and Analysis

We first introduce usage datasets in this paper. Afterward, we evaluate the predictive ability of our proposed models and compare the performance between PPM and LPM models. Additionally, the robustness test is implemented by accessing the performance on several other datasets and comparing it with state-of-art researches. Finally, we employ the prediction results to hedging binary options to demonstrate the effectiveness from the perspective of buyers and sellers.

In this paper, we utilize the **sklearn** library in python to implement the machine learning algorithms [48], and the default parameter settings are used in this paper. For example, MLR adopts default parameters, and RFR contains 500 estimators with bootstrap and criterion of mean squared error, and SVR is with scaled gamma, the kernel of radial basis function, and maximum iteration of $1,000$, and MLP is three hidden-layers perceptrons with 12-4-1 neurons (5-4-1 neurons for LPM) with a solver of adam, a learning rate of 0.001, and iteration of 200.

A. Data Usage

In this paper, we utilize hourly weather data provided by the Central Weather Bureau of Taiwan[1] as the simulation of IoT-based data from solar-power plant and to construct the predictive models. The dataset includes weather records of 608 locations in Taiwan, but only 30 large-scale meteorological stations provide the records of solar-radiation. We randomly select 5 datasets from 30 large-scale meteorological datasets for training and validation, which come from weather stations in different cities, covering the wild range of latitude and longitude in Taiwan. Those datasets are then listed in Table III and called **Datasets A** in the designed predictive model. The first 80% data of **Datasets A** are used for training, and the last 20% data are used for testing.

Additionally, another 5 datasets are used to verify the effectiveness and robustness of the models, which are listed as the testing sample shown in Table IV and called **Datasets B**. The first 80% data of **Datasets B** are used for fine-tune (re-train), and the last 20% data are used for testing. In this paper, we utilize the data from July 2010 to June 2020, and the training period of both **Datasets A** and **B** is from July 2010 to June 2018, and the testing period is from July 2018 to June 2020.

TABLE III: Latitude and longitude of the meteorological stations of **Datasets A**

Station ID	**Latitude (N)**	**Longitude (E)**	**City**
467480	23°50’	120°43’	Chiayi City
467060	24°60’	121°86’	Yilan County
466900	25°16’	121°45’	New Taipei City
467440	22°57’	120°32’	Kaohsiung City
466940	25°13’	121°74’	Keelung City

TABLE IV: Latitude and longitude of the meteorological stations of **Datasets B**

Station ID	**Latitude (N)**	**Longitude (E)**	**City**
467410	22°99’	120°20’	Tainan City
467650	23°88’	120°91’	Nantou County
467540	22°36’	120°90’	Taitung County
466920	25°04’	121°51’	Taipei City
467770	24°26’	120°52’	Taichung City

B. Solar-Radiation Prediction

In this section, we demonstrate the prediction results of the developed PPM and LPM models on **Datasets A** during the testing period, as shown in Table V. Two performance measures are utilized, including the R-squared and the correlation coefficient (C.C.) between the predicted results and the ground truth. Among all the tables in this section, each row represents the performance of each utilized machine learning algorithm on the five weather stations (five columns). In Table V, the values shown in bold are the best performance of the station among all machine learning algorithms.

[1]Central Weather Bureau Observation Data Inquire System, https://e-service.cwb.gov.tw/HistoryDataQuery/index.jsp

TABLE V: Testing performance of prediction models on **Datasets A**

Station	**467480**		**467060**		**466900**		**467440**		**466940**	
Indicators	R-squared	C.C.	R-squared	C.C.	R-squared	C.C.	R-squared	C.C.	R-squared	C.C.
PPM-MLR	0.809	0.899	0.685	0.827	0.704	0.839	0.820	0.906	0.676	0.823
PPM-RFR	**0.827**	**0.910**	**0.726**	**0.852**	**0.748**	**0.865**	**0.841**	**0.917**	**0.729**	**0.854**
PPM-SVR	0.824	**0.910**	0.705	0.843	0.735	0.860	0.829	0.916	0.712	0.846
PPM-MLP	0.817	0.906	0.685	0.828	0.729	0.854	0.830	0.911	0.704	0.840
Station	**467480**		**467060**		**466900**		**467440**		**466940**	
Indicators	R-squared	C.C.	R-squared	C.C.	R-squared	C.C.	R-squared	C.C.	R-squared	C.C.
LPM-MLR	0.806	0.898	0.679	0.824	0.700	0.837	0.817	0.904	0.665	0.816
LPM-RFR	0.810	0.900	0.678	0.824	0.705	0.840	0.828	0.910	0.666	0.816
LPM-SVR	0.809	**0.904**	0.659	0.817	0.697	0.838	0.817	**0.911**	0.646	0.807
LPM-MLP	**0.816**	0.903	**0.708**	**0.843**	**0.710**	**0.843**	**0.829**	0.910	**0.668**	**0.818**

TABLE VI: Testing performance of prediction models on **Datasets B** (robustness test)

Station	**467410**		**467650**		**467540**		**466920**		**467770**	
Indicators	R-squared	C.C.	R-squared	C.C.	R-squared	C.C.	R-squared	C.C.	R-squared	C.C.
PPM-MLR	0.823	0.907	0.706	0.840	0.757	0.870	0.665	0.815	0.808	0.899
PPM-RFR	**0.842**	**0.918**	**0.746**	**0.864**	**0.784**	**0.885**	**0.711**	**0.843**	**0.831**	**0.912**
PPM-SVR	0.836	**0.918**	0.733	0.856	0.767	0.880	0.698	0.837	0.823	0.911
PPM-MLP	0.831	0.912	0.734	0.857	0.771	0.878	0.677	0.824	0.821	0.906
Station	**467410**		**467650**		**467540**		**466920**		**467770**	
Indicators	R-squared	C.C.	R-squared	C.C.	R-squared	C.C.	R-squared	C.C.	R-squared	C.C.
LPM-MLR	0.821	0.906	0.703	0.839	0.753	0.868	0.655	0.809	0.806	0.898
LPM-RFR	0.827	0.910	0.713	0.844	0.760	0.872	0.657	0.811	0.814	0.902
LPM-SVR	0.823	**0.913**	0.710	0.844	0.742	0.869	0.650	0.808	0.806	0.903
PPM-MLP	**0.832**	**0.913**	**0.720**	**0.848**	**0.768**	**0.876**	**0.670**	**0.818**	**0.817**	**0.904**

The first main row of Table V presents the results of the PPM among four machine learning algorithms. In this experiment, the PPM-RFR models obtain the best performance with the highest R-squared and C.C among all machine learning algorithms in **Datasets A**. The best performance of PPM can reach an R-squared of 0.841 and C.C of 0.917. The second main row of Table V presents the results of the LPM. Surprisingly, the results of LPM are quite close to the results of PPM with minor weaknesses. Compared with PPM, the R-square of LPM is reduced by about 0.021, and the C.C is reduced by about 0.012, however, less than half of the features are required for LPM to obtain excellent results. Among various machine learning algorithms utilized in LPM, LPM-RFR and LPM-MLP obtain the best performance, especially for LPM-MLP with a minor advantage. It can be concluded that the historical solar-radiation of the past five days is the important and explanatory feature of predation.

In addition, Table VII shows the average computation time of LPM utilized in MLR, RFR, SVR, and MLP is respectively 47.3%, 54.4%, 41.8%, and 44.7% of the generic PPM. In addition, MLR spends less computational time in all machine learning models, as shown in bold in Table VII.

TABLE VII: Comparing the computation time of PPM and LPM

Model	**Computation Time (ms)**
PPM-MLR	**0.484**
PPM-RFR	2.081
PPM-SVR	9.946
PPM-MLP	21.409
LPM-MLR	**0.229**
LPM-RFR	1.132
LPM-SVR	4.160
PPM-MLP	9.580

In summary, the PPM can obtain precise prediction results with much computation time. As for LPM, it uses less than half of the features and computation time of PPM, and its predictive performance is pretty close to PPM with minor weaknesses. Therefore, we suggest that the best model in the proposed system is LPM-RFR, which meets the IoT-based data-driven scenario and achieves outstanding prediction accuracy with relatively lower computation time (then LPM-MLP).

TABLE VIII: Cumulative payoff and Sharpe ratio for the broker (seller-side)

Station	467410		467650		467540		466920		467770	
Indicators	Payoff	Sharpe	Payoff	Sharpe	Payoff	Sharpe	Payoff	Sharpe	Payoff	Sharpe
T=10	**259.1**	**3.354**	**135.4**	**2.091**	**83.8**	1.050	-12.3	-0.213	**92.5**	1.216
T=15	147.9	2.667	63.1	1.705	-8.9	-0.174	**28.8**	0.942	82.2	1.926
T=20	35.3	1.235	8.2	0.473	50.7	**1.857**	16.2	**0.969**	47.0	**1.988**

TABLE IX: Win rate and average odds for the investors (buyer-side)

Station	467410		467650		467540		466920		467770	
Indicators	$\bar{WR}$	$\bar{ODDS}$	$\bar{WR}$	$\bar{ODDS}$	$\bar{WR}$	$\bar{ODDS}$	$\bar{WR}$	$\bar{ODDS}$	$\bar{WR}$	$\bar{ODDS}$
T=10	0.121	5.351	0.245	3.322	0.211	4.196	0.403	2.525	0.223	3.911
T=15	0.327	2.436	0.534	1.710	0.471	2.148	0.645	1.489	0.490	1.810
T=20	0.684	1.392	0.844	1.172	0.685	1.359	0.851	1.149	0.740	1.265

TABLE X: Comparison with the state-of-the-art approaches

Model	PPM	LPM	SVM-SPP	SVM-AMVs	DDBM
R^2	**0.827**	0.810	-	0.731	-
C.C	**0.910**	0.900	-	-	0.895
MAE	**0.189**	0.197	0.329	-	0.199

C. Robustness Test of Predictive Models

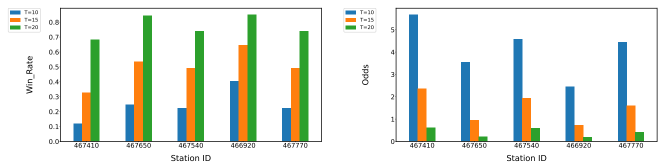
We also perform a robustness test for PPM and LPM conducted on **Datasets B**, and the results are shown in Table VI. Without adjusting the hyper-parameters of models, we only fine-tune (re-train) the machine learning algorithms through the training period of the **Datasets B**, and evaluate the machine learning algorithms through the testing period of the **Datasets B**. In Table VI, the values shown in bold are the best performance of the station among all machine learning algorithms.

The first main row of Table VI presents the results of the PPM. The PPM-RFR models obtain the best performance, that receives the highest R-squared and C.C among all machine learning algorithms in **Datasets B**. The best performance of PPM can reach an R-squared of 0.842 and C.C of 0.918. The second main row of Table VI presents the results of the LPM. Similarly, the results of LPM are quite close to the results of PPM with minor weaknesses, and the LPM-RFR and LPM-MLP obtain the best performance, especially for LPM-MLP with a minor advantage. Compared with PPM, the R-square of LPM is reduced by about 0.016, and the C.C is reduced by about 0.009 in **Datasets B**. However, for the two compared predictive models, less than half of the features are required for LPM. Thus, for the developed LPM, LPM-RFR and LPM-MLP can still obtain the best performance with 0.832 R-squared and 0.913 C.C.

Several state-of-the-art approaches [32], [33], [34] are then compared with the designed two models (PPM and LPM) with the RFR algorithm in terms of the prediction results on station 467480, which can be observed in Table X. The compared approaches include SVM-based short-term solar-power prediction (SVM-SPP) [32], SVM-based model with real-time atmospheric motion vectors (SVM-AMVs) [33], and data-driven-based model (DDBM) [34]. The detail of each approach is described in Section II-A. The measurements include R-squared (R^2), correlation coefficient (C.C.) and mean absolute error (MAE). Note that the units of solar-radiation are different between the researches (affects MAE), and we converted all of them to MJ/m^2, which is mega (10^6) joule per square meter. Also note that the values shown in bold are the best performance of each indicator among all algorithms.

In Table X, it is obvious to see that the designed PPM-RFR and LPM-RFR (with simple and few meteorological observations) obtained better performance compared to the other state-of-the-art approaches among all indicators. In summary, the experimental results show the robustness of the two predictive PPM and LPM in terms of various datasets and measurements.

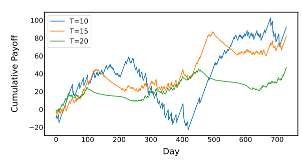
D. Effectiveness of Hedging Option



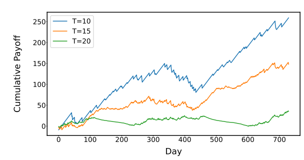
(a) The win rate for investors (b) The odds for investors

Fig. 3: The win rate and odds for investors with different T

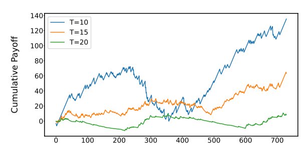
In this section, we evaluate the effectiveness of the proposed hedging options on the **Datasets B** and from the perspective of the broker and the investors (seller-side and buyer-side). We observe the payoff arisen on the broker, and observe the win rate and average odds on the investor.



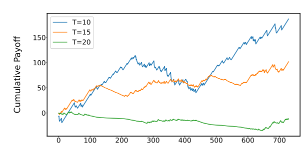
(a) Station 467410



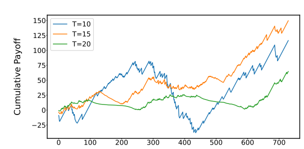
(b) Station 467650



(c) Station 467540



(d) Station 466920



(e) Station 467770

Fig. 4: The cumulative payoff of options for the broker

The predicted solar-radiation described in Section III-C are utilized to obtain the win rate (*WR*) and the odds (*ODDS*) of each hedging option (under $T = 10, 15, 20$ $\frac{MJ}{m^2}$ and commission of 5%). To eliminate outliers, we limit the *ODDS* of hedging options to $[1, 10]$. The reason to set this interval is that if the *ODDS* is greater than 10, it will bring great risks to the broker, and the broker is unwilling to offer such an option. If the *ODDS* is smaller than 1, it always let investors lose money on the hedging option, and the investors are unwilling to buy such an option. To facilitate understanding and implementation, we rule the hedging options to bet on the cumulated solar-radiation within a day (the time unit in this section), and the predictive models will accumulate the hourly predicted solar-radiation of the day as the prediction.

In this experiment, we simulated the proposed hedging system based on the real weather data from July 2018 to June 2020, mentioned in IV-A. Suppose an investor spends one dollar a day on a hedging option (with the same T) without losing generality. The hedging options are provided by a broker, who provides the *ODDS* for each option through the proposed system. From the investor's point of view, he spends one dollar a day on a hedging option with a given *ODDS*. We observe the average value of the received *ODDS* and the win rate (*WR*, probability of low-radiation and that he receives payoff) from the weather data, as shown in Table IX and Fig. 3. From the broker's point of view, he provides the *ODDS* for each option through the proposed system and historical weather data. We observe his income (one dollar a day) and the money he pays under the low-radiation, and his total payoff and the simulated payoff curves can be obtained, as shown in Table VIII and Fig. 4.

Table VIII lists the cumulative payoffs and Sharpe ratio for the broker, where the Sharpe ratio is a financial indicator measuring the trade-off between profitability and risk [49] (higher is better). Note that the values shown in bold are the best performance among all thresholds (rows). From Table VIII, we can observe that the broker can obtain positive payoffs in almost all options, except for T=15 in station 467540 and $T = 10$ in station 466920. Besides, the payoffs can reach a Sharpe ratio of 3.354, which is significantly higher than the Sharpe ratio in the stock market. Fig. 4 shows the curves of cumulative payoffs in different stations. It can be found that the curves usually steadily grow (contributed from the commission as fixed income), and there are several drawdowns due to the low solar radiation. These results show that the profitability for the broker can be obtained, which improves the broker's willingness and motivation to join our novel hedging system for earning fixed income.

We also provide statistical results in the case of low solar-radiation, which are the win rate ($\widehat{WR}$) and average odds ($\widehat{ODDS}$) for investor (buyer-side), as shown in Table IX and Fig. 3. In Fig. 3, it can be found that as the increase of the solar-radiation threshold T, $\widehat{WR}$ increases while $\widehat{ODDS}$ decreases, which makes common sense. Overall, investors receive hedging services from the designed binary options, which requires a certain cost that makes the negative expected return for an investor. For example, in Table IX, an investor near the station 467410 spends 1 dollar a day for hedging option with T of 10 $\frac{MJ}{\mathrm{m}^2}$. The win rate of the investor is 12.1% with averagely odds of 5.351, and the expected return is negative and is $12.1\% \times 5.351 - 1 = -0.353$. However, when the solar-radiation is less than 10 MJ/m^2, they can averagely earn 5.351 dollars to cover the loss from solar-power generation, which is the goal of hedging services. The buyer-side experimental results in Table IX show the effectiveness of the proposed hedging system. The hedging system costs little every day and provides a large payoff when the solar-radiation is low to make up for the loss of solar-power investment.

V. Conclusions

With high concern for environmental protection in recent years, a mature solar business model has been established. However, some uncertainties are left without solutions, such as the risk of low solar-radiation which may result in investment losses. To address these issues, we propose a novel hedging system utilized in solar-power business, which adopts the edge-based predictive models with IoT data for solar radiation, and the hedging binary option. The solar panels and weather sensors are collected through IoT-based data in the proposed system, and edge-based models are constructed for predicting solar-radiation. The precise prediction model (PPM) is proposed to predict solar-radiation in high precision, and the light predictive model (LPM) is proposed to meet the low latency and less computational cost on the edge-system.

Our experimental results indicate that random forest regression achieved the best performance, and PPM and LPM with random forest regression obtain R-squared of 0.841 and 0.828 and correlation coefficient of 0.917 and 0.910, respectively. Besides, LPM uses about half of the computation time to obtain accuracy similar to PPM. Compared with state-of-the-art models, our models have achieved better performance in all metrics, which demonstrates the robustness and effectiveness of the proposed models.

As for the proposed hedging options, our simulation results show that the broker can obtain stable payoffs with the highest Sharpe ratio of 3.354. Regarding the investor's point of view, the hedging system costs little financial overhead and provides a large payoff (when low solar-radiation) to make up for the loss of solar-power investment. Our simulation results show the effectiveness of the proposed hedging system for investors (buyer-side), simultaneously, present the motivation of the broker (seller-side) to join our system to earn a fixed income.

In the future, we plan to enhance prediction with various features (such as weather forecasts, satellite, and radar images). Besides, the designed hedging options will be extended to both buyer-side and seller-side for investors (investors can also sell options to solar companies and other investors) to increase the flexibility of the hedging system.

Acknowledgments

This research is partially supported by the National Centre for Research and Development under the project Automated Guided Vehicles integrated with Collaborative Robots for Smart Industry Perspective and the Project Contract no. is: NOR/POLNOR/CoBotAGV/0027/2019 -00, and by the Ministry of Science and 682 Technology (MOST), Taiwan, under the grant code of MOST 109-2221-E-027-106.

References

[1] K. P. Shine, J. S. Fuglestvedt, K. Hailemariam, and N. Stuber, "Alternatives to the global warming potential for comparing climate impacts of emissions of greenhouse gases," *Climatic Change*, vol. 68, no. 3, pp. 281–302, 2005.

[2] P. Collarile, E. Bidoli, F. Barbone, L. Zanier, S. Del Zotto, S. Fuser, F. Stel, C. Panato, I. Gallai, and D. Serraino, "Residence in proximity of a coal-oil-fired thermal power plant and risk of lung and bladder cancer in north-eastern italy. a population-based study: 1995–2009," *International Journal of Environmental Research and Public Health*, vol. 14, no. 8, pp. 860–870, 2017.

[3] O. Ellabban, H. Abu Rub, and F. Blaabjerg, "Renewable energy resources: Current status, future prospects and their enabling technology," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 748–764, 2014.

[4] T. Ackermann, *Wind power in power systems*. John Wiley & Sons, 2005.

[5] G. K. Singh, "Solar power generation by pv (photovoltaic) technology: A review," *Energy*, vol. 53, pp. 1–13, 2013.

[6] Y. Lei, X. Lu, M. Shi, L. Wang, H. Lv, S. Chen, C. Hu, Q. Yu, and S. D. H. da Silveira, "Swot analysis for the development of photovoltaic solar power in africa in comparison with china," *Environmental Impact Assessment Review*, vol. 77, pp. 122–127, 2019.

[7] J. C. W. Lin, G. Srivastava, Y. Zhang, Y. Djenouri, and M. Aloqaily, "Privacy preserving multi-objective sanitization model in 6g iot environments," *IEEE Internet of Things Journal*, 2020.

[8] G. Fortino, W. Russo, C. Savaglio, W. Shen, and M. Zhou, "Agent-oriented cooperative smart objects: From iot system design to implementation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 11, pp. 1939–1956, 2017.

[9] G. Aloi, G. Caliciuri, G. Fortino, R. Gravina, P. Pace, W. Russo, and C. Savaglio, "Enabling iot interoperability through opportunistic smartphone-based mobile gateways," *Journal of Network and Computer Applications*, vol. 81, pp. 74–84, 2017.

[10] C. Savaglio, P. Gerace, G. Di Fatta, and G. Fortino, "Data mining at the iot edge," in *International Conference on Computer Communication and Networks*, 2019, pp. 1–6.

[11] M. Baza, N. Lasla, M. Mahmoud, G. Srivastava, and M. Abdallah, "B-ride: Ride sharing with privacy-preservation, trust and fair payment atop public blockchain," *IEEE Transactions on Network Science and Engineering*, 2019.

[12] W. Al Amiri, M. Baza, K. Banawan, M. Mahmoud, and W. Alasmary, "Privacy-preserving smart parking system using blockchain and private information retrieval," in *International Conference on Smart Applications, Communications and Networking*, 2019, pp. 1–6.

[13] D. J. Hand and N. M. Adams, "Data mining," *Wiley StatsRef: Statistics Reference Online*, pp. 1–7, 2014.

[14] D. T. Larose and C. D. Larose, *Discovering knowledge in data: An introduction to data mining*. John Wiley & Sons, 2014, vol. 4.

[15] X. Wu, X. Zhu, G. Q. Wu, and W. Ding, "Data mining with big data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 97–107, 2013.

[16] E. Alpaydin, *Introduction to machine learning*. MIT press, 2020.

[17] J. C. W. Lin, Y. Shao, Y. Djenouri, and U. Yun, "Asrnn: A recurrent neural network with an attention model for sequence labeling," *Knowledge-Based Systems*, 2020.

[18] W. Yu, F. Liang, X. He, W. G. Hatcher, C. Lu, J. Lin, and X. Yang, "A survey on the edge computing for the internet of things," *IEEE Access*, vol. 6, pp. 6900–6919, 2017.

[19] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.

[20] M. Chen, W. Li, G. Fortino, Y. Hao, L. Hu, and I. Humar, "A dynamic service migration mechanism in edge cognitive computing," *ACM Transactions on Internet Technology*, vol. 19, no. 2, pp. 1–15, 2019.

[21] M. S. Mekala and P. Viswanathan, "Energy-efficient virtual machine selection based on resource ranking and utilization factor approach in cloud computing for iot," *Computers & Electrical Engineering*, vol. 73, pp. 227–244, 2019.

[22] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[23] R. Casadei, G. Fortino, D. Pianini, W. Russo, C. Savaglio, and M. Viroli, "Modelling and simulation of opportunistic iot services with aggregate computing," *Future Generation Computer Systems*, vol. 91, pp. 252–262, 2019.

[24] P. Pace, G. Aloi, R. Gravina, G. Caliciuri, G. Fortino, and A. Liotta, "An edge-based architecture to support efficient applications for healthcare industry 4.0," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 481–489, 2018.

[25] M. S. Mekala and P. Viswanathan, "A survey: Smart agriculture iot with cloud computing," in *International Conference on Microelectronic Devices, Circuits and Systems*, 2017, pp. 1–7.

[26] R. E. Hoyt and H. Khang, "On the demand for corporate property insurance," *Journal of Risk and Insurance*, pp. 91–107, 2000.

[27] A. Goetzberger and V. U. Hoffmann, *Photovoltaic solar energy generation*. Springer Science & Business Media, 2005, vol. 112.

[28] S. Murmson. (2018) The output watts of solar panels. [Online]. Available: https://sciencing.com/output-watts-solar-panels-6946.html

[29] M. Bernardi, N. Ferralis, J. H. Wan, R. Villalon, and J. C. Grossman, "Solar energy generation in three dimensions," *Energy & Environmental Science*, vol. 5, no. 5, pp. 6880–6884, 2012.

[30] L. Partain, R. Hansen, S. Hansen, D. Bennett, A. Newlands, and L. Fraas, "Swanson's law plan to mitigate global climate change," in *IEEE 43rd Photovoltaic Specialists Conference*, 2016, pp. 3335–3340.

[31] R. M. Swanson, "A vision for crystalline silicon photovoltaics," *Progress in Photovoltaics: Research and Applications*, vol. 14, no. 5, pp. 443–453, 2006.

[32] J. Zeng and W. Qiao, "Short-term solar power prediction using a support vector machine," *Renewable Energy*, vol. 52, pp. 118–127, 2013.

[33] H. S. Jang, K. Y. Bae, H. S. Park, and D. K. Sung, "Solar power prediction based on satellite images and support vector machine," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1255–1263, 2016.

[34] H. Long, Z. Zhang, and Y. Su, "Analysis of daily solar power prediction with data-driven approaches," *Applied Energy*, vol. 126, pp. 29–37, 2014.

[35] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*. MIT press, 2018.

[36] J. C. W. Lin, Y. Shao, Y. Djenouri, and U. Yun, "Asrnn: a recurrent neural network with an attention model for sequence labeling," *Knowledge-Based Systems*, vol. 212, no. 5, p. 106548, 2021.

[37] F. Khademi, S. M. Jamal, N. Deshpande, and S. Londhe, "Predicting strength of recycled aggregate concrete using artificial neural network, adaptive neuro-fuzzy inference system and multiple linear regression," *International Journal of Sustainable Built Environment*, vol. 5, no. 2, pp. 355–369, 2016.

[38] G. Forkuor, O. K. Hounkpatin, G. Welp, and M. Thiel, "High resolution mapping of soil properties using remote sensing variables in south-western burkina faso: A comparison of machine learning and multiple linear regression models," *PloS One*, vol. 12, no. 1, p. e0170478, 2017.

[39] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[40] Y. Kaneda and H. Mineno, "Sliding window-based support vector regression for predicting micrometeorological data," *Expert Systems with Applications*, vol. 59, pp. 217–225, 2016.

[41] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[42] B. Babar, L. T. Luppino, T. Boström, and S. N. Anfinsen, "Random forest regression for improved mapping of solar irradiance at high latitudes," *Solar Energy*, vol. 198, pp. 81–92, 2020.

[43] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005.

[44] L. Noriega, "Multilayer perceptron tutorial," *School of Computing, Staffordshire University*, 2005.

[45] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2.

[46] A. Nekritin, *Binary Options: Strategies for Directional and Volatility Trading*. John Wiley & Sons, 2012, vol. 595.

[47] H. Raw, *Binary options: Fixed odds financial bets*. Harriman House Limited, 2011.

[48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[49] W. F. Sharpe, "The sharpe ratio," *Journal of Portfolio Management*, vol. 21, no. 1, pp. 49–58, 1994.