Journal of Theoretical and Applied Information Technology
15th January 2022. Vol.100. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ASSOCIATION RULE ALGORITHM FOR WEATHER FACTORS ESTIMATION USING ON-GRID SOLAR PV POWER SYSTEM

[1,2]MOHAMMAD A OBEIDAT, [3]AYMAN M MANSOUR*, [4]FAYEZ KHAZALAH, [2]JALAL ABDALLAH

1 Department of Electrical Engineering, Faculty Engineering, Al-Ahliyya Amman University, Amman 19328, Jordan

2 Department of Electrical Power and Mechatronics Engineering, College of Engineering, Tafila Technical University, Tafila 66110, Jordan

3 Department of Communications, Electronics, and Computer Engineering, College of Engineering, Tafila Technical University, Tafila 66110, Jordan

4 Department of Information Systems, Al al-Bayt University, Mafraq, Jordan

*Correspondence: mansour@ttu.edu.jo

ABSTRACT

Many stations are established worldwide to measure the weather factors such as temperature, humidity, wind speed, and cloud cover. In this paper, A developed model using association rule mining algorithm is utilized to predict weather factors from real power photovoltaic PV system on-grid. This model is tested on Tafila Technical university PV system data. The developed model contributes successfully as a virtual weather station to forecast the weather in Tafila-Jordan region. The extracted rules of the model are examined by a weather forecasting expert in order to validate the developed system. The results show high confidence and accuracy. The predicted weather factors from PV power values matched approximately 97% of the results given by the experts.

Keywords: *Association Rule; Power; Photovoltaic; Weather Conditions; Estimation.*

1. INTRODUCTION

The field of artificial intelligence is getting more and more attention in the recent years. Data mining techniques have been successfully applied in many applications such as: power applications [1], medical applications and diagnoses as in [2-6], and [7], communications [8], facilitate user tasks such as [9-12], and [13] rainfall forecasting and predicting [14]. In these applications, different machine learning techniques are used such decision tree [1,5], reconstruction phase space [6], fuzzy logic [3,7-9], association rules [10], multi-agent system [4,11], naïve base [12], Neural network [2,13,14].

One of the well-known data mining techniques is association rule mining. It was introduced first in [15] to extract significant association rules between sets of items from customer transactions (aka, basket data) of a very big retailing company. The proposed algorithm is mostly known as AIS afterwards. The discovered rule is composed of two parts: antecedent and consequent, and only rules with one-item consequent were considered.

An example of a mined rule might be that “85% of all transactions where a customer buys bread and eggs, also buys milk.” The list {bread, eggs, milk} is called itemset, whereas the antecedent is {bread, eggs} and the consequent is {milk}. 90% represents the confidence in this rule, which measures its strength. The confidence metric is simply the conditional probability that the consequent will occur given the antecedent, or by the mathematical notation is P(consequent antecedent). Support is another metric that is used to evaluate the rule. It represents the percentage of the rule’s frequent itemset to the total number of transactions. Thus, a rule with high support means that this rule is much significant compared to a one with low support. A support threshold is usually defined so we consider only the itemsets that satisfy this threshold. In [16], the database of transactions is scanned many times.

In the first scan, candidate 1-itemsets along with their support count are discovered. The candidate itemsets with support count less than the support threshold are removed, and the remaining ones represent the frequent 1-itemsets, which are used as an input to the next database scan to compute candidate 2-itemsets. This process continues until no more candidates or frequent itemsets are left. However, this algorithm has serious drawbacks [16]. First, it needs to scan the database many times. In addition, it generates many needless candidate itemsets with their support count that are removed later due to their low support, which waists resources and degrades the algorithm's performance.

In [17], the same authors presented a new improved algorithm to overcome the drawbacks of AIS, and called it the Apriori algorithm. The results showed that the new algorithm has better performance over AIS. Afterwards, it became the mostly used algorithm for mining association rules. It differs from AIS by using a different method for generating candidate itemsets and new pruning techniques. In addition, the author has allowed for the consequent of an association rule to have more than one item, making the rules more general as compared to that introduced in [15]. The popularity of the Apriori algorithm comes from its primary Apriori property and support-based pruning [18]. The former states, "if an itemset is frequent, then all of its subsets must also be frequent." The latter states, "if an itemset is infrequent, then all of its supersets are infrequent." Thus, during the generation process of frequent itemsets, all supersets of infrequent itemsets are pruned. However, the Apriori algorithm still has serious issues, mainly in generating candidate itemsets. Many studies have been proposed to improve the performance and efficiency of the Apriori algorithm.

[19] proposed a frequent pattern tree structure (FP-tree) which is a small compressed tree-based version of the original database. It then uses an FP-tree based mining method, called FP-growth, to mine frequent patterns using pattern fragment growth, thus avoiding the high cost of candidate itemsets generation process. In addition, the use of this tree structure avoids scanning the database repeatedly. It needs only two database scans to generate the frequent itemsets.

Another similar tree-based algorithm called RARM (Rapid Association Rule Mining) was proposed in [20]. It uses a trie-like tree structure that stores the support of 1-itemset and 2-itemsets in each transaction of the database. Thus, it can extract 1-itemsets and 2-itemsets without scanning the database, which reduces the exponential search of generating 2-itemsets that the Apriori suffers from. However, it uses the Apriori algorithm to extract the remaining larger itemsets.

To resolve the performance bottleneck of 2-itemsets generation in the Apriori algorithm, [21] proposed a hash-based algorithm called DHP (Direct Hashing and Pruning). Its main goal is to generate a very small candidate 2-itemsets compared to the original algorithm. Thus, it trims a large portion of the database early and reduces the cost of later passes over the database and cuts the memory and storage usage significantly. Association rule mining is used increasingly in a wide-area of applications. In addition to market basket analysis, its applications, for examples, include medical diagnoses, protein sequences, and census data [22].Recently, association rule mining is used in power systems to predict problems early, which enables operators or power systems from taking urgent actions in anticipation of these problems, or improving operation, management and maintaining of power systems. Next, we provide some examples on using association rules in power systems.

Geomagnetically induced currents (GICs) that result from a solar storm or geomagnetic disturbance (GMD) can severely affect power grids. They may damage power grid transformers, cause failure of their components, like circuit breakers and relays, and leads to large-scale power outages. To see which physical variables, related to GMDs are helpful to predict the effects of GMDs on power grid, [23] applied association rule mining on these physical variables. Based on the results, manual or automatic mitigating actions can be taken.

To know the fault causes and locations in the transmission lines of power grids, [24] applied Apriori algorithm on data related to power grids. The outcome of mining association rules is then used by grid operators for issuing early warnings and/or taking preventive actions in advance. [25] used association rule mining to improve the monitoring application that utilizes data collected from power devices. It uses an improved version of the Apriori algorithm based on Upper Triangular itemsets Matrix (UTM) to mine association rules from relevant data. In [26], association rule mining is applied on defect data collected from Relay Protection Devices (RPDs) to discover defect patterns of RPDs, thereby improving operation managing, and maintaining of RPDs.

In this paper association rule mining is used to extract the rules that relate the generated amount of power with weather factors. The proposed algorithm is suitable to be used in areas where weather stations are not available. It helps in cost saving and enhances the overall weather prediction quality.

2. TAFILA TECHNICAL UNIVERSITY PV STATION

The objective of this project is to electrify the available electrical Loads in AL-Tafila Technical University (TTU) located in Tafila using an on-grid PV system with a total peak power of 1MW. In this project, the electricity demand in the university buildings is met by the solar PV system energy, after being converted from DC to AC and the excess is directly fed to the grid

The Solar PV system consists mainly of multi crystalline Photovoltaic modules which will be the source for electrifying the loads with the support of the national Grid Company (On Grid System), on-grid high efficiency inverters, monitoring system, and the necessary PV mounting structures. Figure 1 shows Tafila Technical university system site view.

Figure 1: Tafila Technical University System Site-View

To accomplish this task, we are proposing a comprehensive turn-key solution including the design, sizing, supply, implementation, and commissioning for this 1 MWp on-grid system. Tafila Technical University in Tafila which is situated at an altitude of 1252 m above sea level and at latitude of 30.20 north of the equator has a high potential of solar energy.

The annual daily average of solar radiation at the site received at a 100 tilt angle is 5.9 kWh/m^2. The distribution of the monthly average of solar radiation intensity received on a horizontal plane and that received on the suggested tilt angles is illustrated in Table 1.

Table 1: Meteorological Data Of The Site

Month	Global Horizontal (kWh/m^2)	Ambient Temp. (°C)
January	101.7	7.68
February	114	8.94
March	165.9	12.51
April	201.7	17.37
May	241.4	21.91
June	253.7	24.78
July	261.5	27.28
August	236.2	26.67
September	199.8	24.18
October	157.1	19.74
November	114.2	13.69
December	93.6	9.57
Year	2140.8	17.91

The PV system for building 1 inside the university as an example can be described through its single line diagram shown in Figure 2.

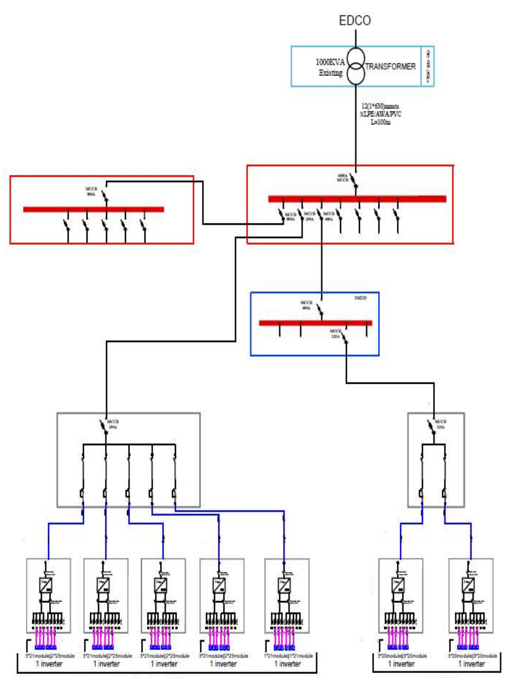

Figure 2: Single Line Diagram Of Building 1

Journal of Theoretical and Applied Information Technology
15th January 2022. Vol.100. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The system consists of a 30.6 kWp PV arrays made up of 120 multi crystalline modules, each with a maximum power of 255 W and an efficiency of 15.7%. These arrays are connected to an on-grid inverter made by ABB company. The output of this inverter is connected to national electric grid via the necessary panel distribution board. Figure 3 shows lay-Out of the Modules on the Roof of Building 1.

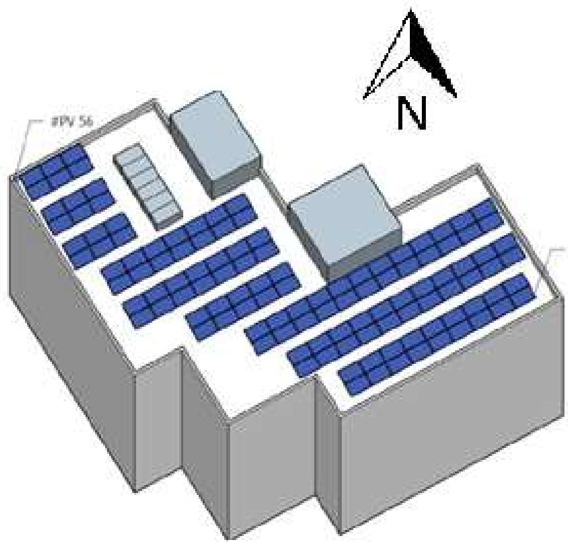

Figure 3: Lay-Out of the Modules On The Roof Of Building 1.

The heart of any PV system is the inverter. The inverter converts the direct current from the PV array to grid-compliant alternating current. The offered inverters are characterized by its 98.2% high efficiency, over voltage protection on input and output sides, water proof design, simplicity, and flexibility. The power capacity and the generated output energy from all the PV systems installed on the roofs of all the buildings are shown in Table 2.

3. ASSOCIATION RULE BASED WEATHER PREDICTION MODEL

3.1 Power and Weather Data Collection Stage

Two databases are used in this paper. The first database is power, which was extracted from the photovoltaic power plant at Tafila Technical University. The data generated by the power plant contains the energy produced from the solar system. For each day, a reading is taken over a period of three years from 01-01-2017 to 31-12-2019.

Table 2: Power Capacity and Energy Yield For All PV Systems Installed On The Roofs On The Existing Buildings

Building Number	**Capacity (kWp)**	**Annual Generated Energy (kWh)**
1	30.6	56517
2	13.005	23839
3	22.95	42469
4	183.6	342372
5	90.27	168240
6	55.08	103223
7	23.46	43866
8	13.77	25437
9	22.44	41947
10	20.4	38065
11	24.48	45487
13	30.6	57105
14	163.71	305438
Hanger 1	57.63	106000
Hanger 2	91.8	161129
Car park	156.06	286164
Total	999.855	1847298

Figure 4 shows a statistical graphical summary of the electrical power data collected over a period of three years. This plot provides a visual summary of the data in order to identify the median value, the dispersion of the data set, and signs of skewness. From whisker plot shown in Figures 4, the statistical data: median, minimum, maximum, first quartile, and third quartile for each month during three years.

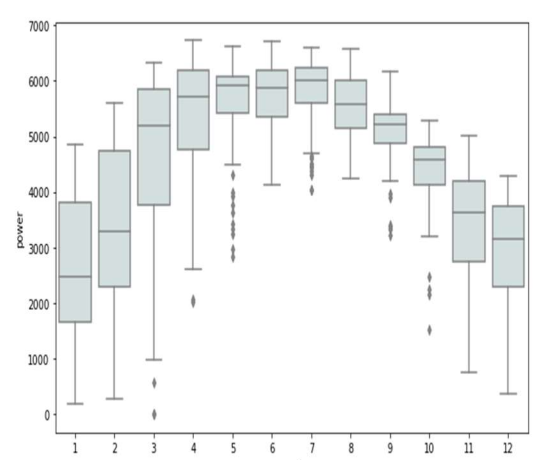

Figure 4: Statistical Graphical Summary Of The Electrical Power Data

Journal of Theoretical and Applied Information Technology
15th January 2022. Vol.100. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The other database is weather data which was extracted from the meteorological station. In weather data, there is a set of weather factors. Statistical graphical summary for each weather factor such as: temperature, cloud cover, humidity, and wind speed is shown in Figure 5, Figure 6, Figure 7, and Figure 8 respectively.

Similar statistical information data such as median, minimum, maximum, first quartile, and third quartile for each month during the period 2009 - 2019 is displayed in Whisker plot.

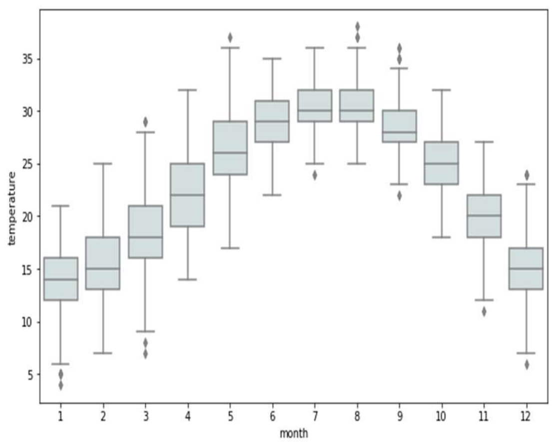

Figure 5: Statistical Graphical Summary Of The Temperature Data

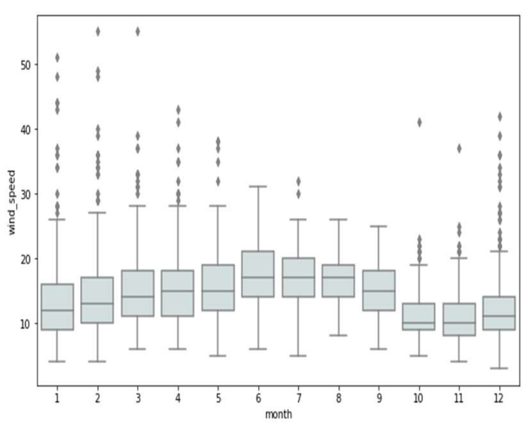

Figure 6: Statistical Graphical Summary Of The Wind-Speed Data

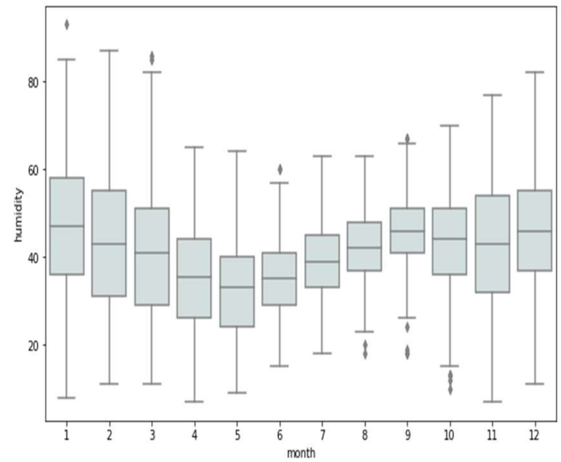

Figure 7: Statistical Graphical Summary Of The Humidity Data

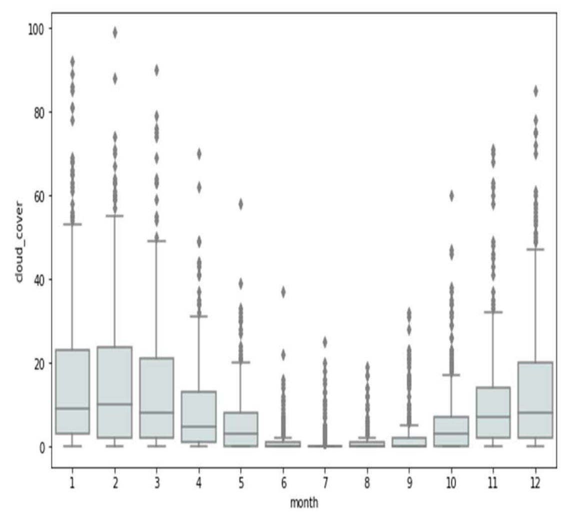

Figure 8: Statistical Graphical Summary Of The Cloud Cover Data

One of the important plots used to give preliminary study of the relations between system parameters, is heat map. Heat map plot is shown in Figure 9 for a set of factors which consists of power, temperature, wind speed, cloud cover, and humidity during three years. Positive, negative, or zero correlation helps in finding initial relations between various factors. The heat map plot encourages authors to apply association rule-based model to find a set of rules that relate these factors to one another. Then take the best of these rules to predict the weather conditions from the power factor.

Journal of Theoretical and Applied Information Technology
15th January 2022. Vol.100. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

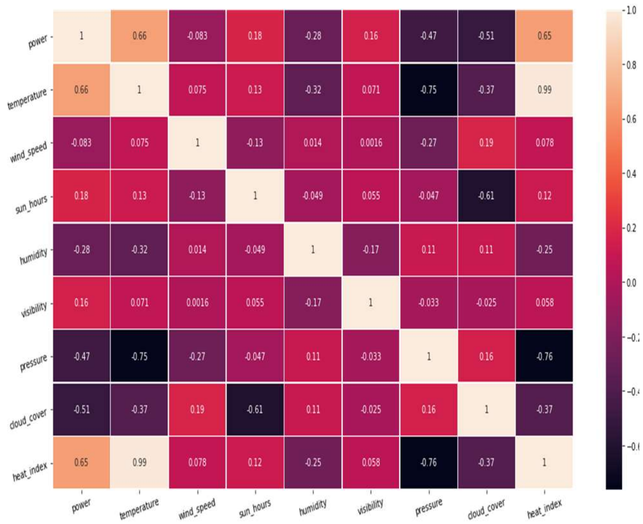

Figure 9: Heat Map Plot Of System Parameters

3.2 Preprocessing Stage

Weather and power databases are combined together in one database as shown in Figure 10. Unnecessary data has been eliminated and four factors are kept. In the preprocessing of the data, the missing values are replaced by taking the average reading in this month and using it instead of the missing value, whether by energy data or weather data.

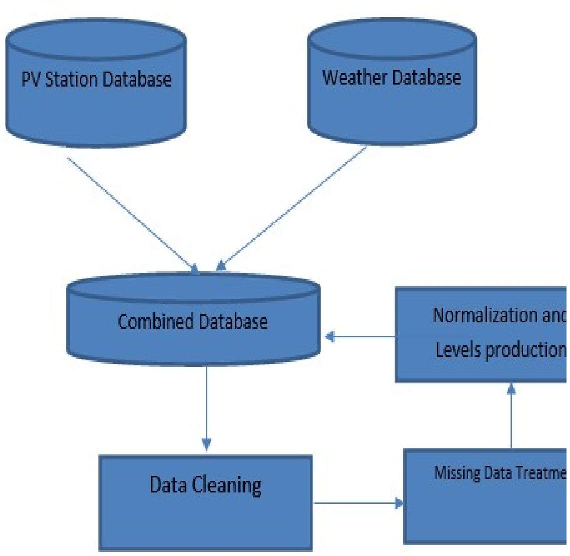

Figure 10: Data Collection And Pre-processing Block Diagram

The second stage of the preprocessing is the normalization as shown in (1)

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Where x is the weather factor variable, x_{norm} is the normalization of x, min(x) is the minimum value of x, and max(x) is the maximum value of x.

The normalization is necessarily in order to convert the values into ranges as shown in the Table 3. This conversion is consistent with the association rule algorithm.

Table 3: Ranges And Levels For System Parameters

Power	Range	0.8-1.0	0.6-0.79	0.4-0.59	0.0-0.39
	Level	Very High	High	Low	Very Low
Temperature	Range	0.85-1.0	0.66-0.84	0.33-0.65	0.0-0.32
	Level	Very High	High	Low	Very Low
Humidity	Range	Humid	Comfort		Dry
	Level	0.6-1.0	0.3-0.59		0.0-0.29
Cloud	Range	Cloudy	Partially Cloudy		Clear
	Level	0.2-1.0	0.1-0.19		0.0-0.18
Wind	Range	0.5-1.0	0.33-0.49		0.0-0.32
	Level	Fast	Medium		Slow

The data sets are normalized, then a level is assigned for each normalized value. A sample of the databases for power, temperature, and wind speed is shown in Table 4, also humidity and cloud cover sample databases are shown in Table 5.

Table 4: Database Example 1

Power		
Power (kW)	**Normalized value**	**Level**
0	0	very low
146.1	0.021	very low
186.9	0.027	very low
194.6	0.028	very low
275.8	0.041	very low
275.9	0.041	very low
282.3	0.041	very low
303.8	0.045	very low
309.8	0.046	very low
332.6	0.049	very low
Temperature		
Temp (°C)	**Normalized value**	**Level**
18	0.833	High
13	0.667	High
15	0.889	Very High
10	0.389	Low
16	0.528	Low
10	0.917	Very High
14	0.639	Low
15	0.806	High
15	0.333	Low
12	0.889	Very High
Wind Speed		
wind speed (Km/h)	**Normalized value**	**Level**
13	0.386	Medium
29	0.114	Slow
44	0.250	Slow
13	0.386	Medium
44	0.250	Slow
21	0.295	Slow
17	0.250	Slow
10	0.341	Medium
13	0.636	Fast
15	0.273	Slow

Table 5: Database Example 2

Humidity		
Humidity (%)	**Normalized value**	**Level**
49	0.314	Comfort
55	0.371	Comfort
50	0.486	Comfort
68	0.971	Humid
52	0.586	Comfort
61	0.500	Comfort
68	0.586	Comfort
54	0.643	Humid
64	0.700	Humid
52	0.500	Comfort
Cloud Cover		
Cloud Cover (%)	**Normalized value**	**Level**
18	0.033	Clear
74	0.637	Cloudy
39	0.000	Clear
70	0.604	Cloudy
69	0.011	Clear
39	0.000	Clear
55	0.011	Clear
6	0.000	Clear
75	0.571	Cloudy
25	0.000	Clear

3.3 Training and Testing Stage

The data was divided into three parts: training, testing, and validation. in the training part, 70% of the data is used to build the rule-based model using Association Rule Algorithm, which uses confidence parameter to check the strongest of each rule in the model. After building the training model, 15% of the data is used to test the model, and the same parameter confidence is used for the most powerful measure of each rule. This is used to be sure that our model can be generalized for the other data in the same manner. Finally, 15% of the data is used to validate the association rules extracted from the first stage, the highest confidence measure of the validation data for each association rule gives it high priority to be one of the best rule in the model. Figure 11 shows the developed model block

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

diagram.

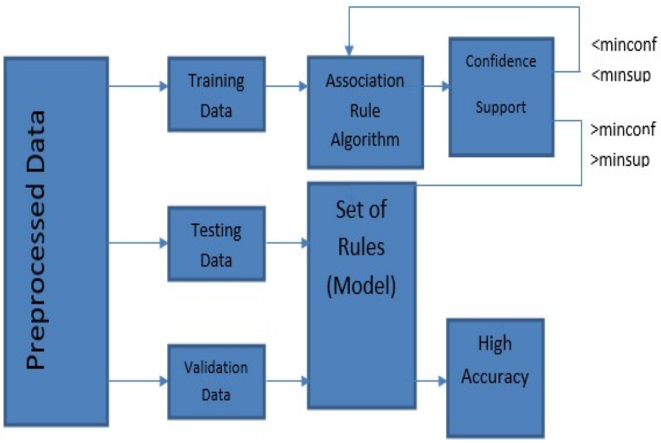

Figure 11. Developed Model Block Diagram

The definition of an association rule algorithm is an implication of the form (X ⇒ Y), where X ⊆ I, Y ⊆ I, and X ∩ Y = ϕ. I is the set of unique items in the database D, X is the antecedent of the rule, and Y is the consequence of the rule. X ⊔ Y is a frequent k-itemset that satisfies minsup. A k-itemset is an itemset that contains k items. An association rule is considered strong if it satisfies both minsup and minconf that represent the minimum support and confidence thresholds, respectively, which are hyper-parameters used by the algorithm to evaluate the importance and strength of mined association rules.

The support is the percentage of an itemset frequency to the total number of records in D. In other words, in how many records of D the itemset appeared. In regarding to an association rule, for example, (X ⇒ Y), the support of (X ⊔ Y) is the number of records that contain both X and Y over the total number of records in D. An itemset is considered a candidate itemset if its support satisfies minsup. The confidence of an association rule, say (X ⇒ Y), is the number of records that contain (X ⊔ Y) to the total number of records that contain X only. It is a metric used to measure the strength of an association rule. Mathematically as (2),

$$confidence\ (X \Rightarrow Y) = \frac{support\ (X \cup Y)}{support\ (Y)} \qquad (2)$$

A strong association rule is the one with a confidence value that satisfies minconf. The problem of mining association rules, from a database D, is a process that composes two steps: 1) Generate frequent or large itemsets 2) generate strong association rules from frequent itemsets.

Association rule algorithm is used to acquire the relations between five factors. The factors are power, temperature, humidity, clouds, and wind speed. The power consists of four levels, Very Low (P_{VL}), Low (P_L), High (P_H), and Very High (P_{VH}). The temperature also consists of four levels: very low (T_{VL}), very low (T_L), high (T_H), and very high (T_{VH}). The humidity consists of three levels: dry (H_D), humid (H_H), and comfort (H_C). The cloudy factor consists of three levels: Cloudy (C_C), partly cloudy (C_P), and clear (C_{CL}). The wind speed consists of three levels: fast (W_F), medium (W_M), and slow (W_S) as shown in Table 6 and Table 7.

Table 6: Humidity, Cloud Over, And Wind Speed Levels

Humidity Levels		
Dry	Humidity	Comfort
(H_D)	(H_H)	(H_C)
Cloud Cover Levels		
Cloudy	Partially Cloud	Clear
(C_C)	(C_P)	(C_{CL})
Wind Speed Levels		
Fast	Medium	Slow
(W_F)	(W_M)	(W_S)

Table 7: Power And Temperature Levels

Power Levels			
Very Low	Low	High	Very High
(P_{VL})	(P_L)	(P_H)	(P_{VH})
Temperature Levels			
Very Low	Low	High	Very High
(T_{VL})	(T_L)	(T_H)	(T_{VH})

In order to proceed in association rules, the following steps are conducted to achieve best rules that describe the hidden relations between the generated power and weather factors:

Step 1: Duo factors array is generated. Each factor level is combined with the other factor level, therefore only one pair of factor levels is created for each element in the array. As an example, power factor level such as (P_H) is combined with all other factors levels such as (T_{VH}), (H_D), (C_P), and (W_F) to produce (P_H T_{VH}), (P_H H_D), (P_H C_P), and (P_H W_F) and not commutative process, so the pair factor levels

Journal of Theoretical and Applied Information Technology
15th January 2022. Vol.100. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(T_{VH} P_H), (H_D P_H), (C_P P_H), and (W_F P_H) are not allowed as shown in Table 8.

Table 8: Duo Factors Array

	P_{VL}	P_L	P_H	P_{VH}
T_{VL}	P_{VL} T_{VL}	P_L T_{VL}	P_H T_{VL}	P_{VH} T_{VL}
T_L	P_{VL} T_L	P_L T_L	P_H T_L	P_{VH} T_L
T_H	P_{VL} T_H	P_L T_H	P_H T_H	P_{VH} T_H
T_{VH}	$P_{VL}T_{VH}$	P_L T_{VH}	P_H T_{VH}	P_{VH} T_{VH}
H_D	P_{VL} H_D	P_L H_D	P_H H_D	P_{VH} H_D
H_C	P_{VL} H_C	P_L H_C	P_H H_C	P_{VH} H_C
H_H	P_{VL} H_H	P_L H_H	P_H H_H	P_{VH} H_H
C_C	P_{VL} C_C	P_L C_C	P_H C_C	P_{VH} C_C
C_P	P_{VL} C_P	P_L C_P	P_H C_P	P_{VH} C_P
C_{CL}	P_{VL} C_{CL}	P_L C_{CL}	P_H C_{CL}	P_{VH} C_{CL}
W_S	P_{VL} W_S	P_L W_S	P_H W_S	P_{VH} W_S
W_M	$P_{VL}W_M$	P_L W_M	P_HW_M	P_{VH} W_M
W_F	P_{VL} W_F	P_L WF	P_H W_F	P_{VH} WF

Step 2: Trio factors array is generated. Each pair of factor levels generated from Duo factors array is combined with the other factor level, therefore only one trio factor levels is created for each element in the array. As an example, power factor level such as (P_H T_{VH}) is combined with all other factors levels such as (H_D), (C_P), and (W_F) to produce (P_H T_{VH} H_D), (P_H T_{VH} C_P), and (P_H T_{VH} W_F) and not commutative process.The trio factor levels (H_D P_H T_{VH}), (C_P P_H T_{VH}), and (W_F P_H T_{VH}) are not allowed as shown in Table 9.

Step 3: Quartet factors array is generated. Each trio factor levels generated from Trio factors array is combined with the other factor level, therefore only one quartet factor levels is created for each element in the array. As an example, power factor level such as (P_H T_{VH} H_D) is combined with all other factors levels such as (C_P), and (W_F) to produce (P_H T_{VH} H_D C_P), and (P_H T_{VH} H_D W_F) and not commutative process, so the trio factor levels (C_P P_H T_{VH} H_D), and (W_F P_H T_{VH} H_D) are not allowed as shown Table 10.

Table 9: Trio Factors Array

	P_{VL} T_{VL}	P_L T_{VL}	P_H T_{VL}	P_{VH} T_{VL}
H_D	P_{VL} T_{VL} H_D	P_L T_{VL} H_D	P_H T_{VL} H_D	P_{VH} T_{VL} H_D
H_C	P_{VL} T_{VL} H_C	P_L T_{VL} H_C	P_H T_{VL} H_C	P_{VH} T_{VL} H_C
H_H	P_{VL} T_{VL} H_H	P_L T_{VL} H_H	P_H T_{VL} H_H	P_{VH} T_{VL} H_H
C_C	P_{VL} T_{VL} C_C	P_L T_{VL} C_C	P_H T_{VL} C_C	P_{VH} T_{VL} C_C
C_P	P_{VL} T_{VL} C_P	P_L T_{VL} C_P	P_H T_{VL} C_P	P_{VH} T_{VL} C_P
C_{CL}	P_{VL} T_{VL} C_{CL}	P_L T_{VL} C_{CL}	P_H T_{VL} C_{CL}	P_{VH} T_{VL} C_{CL}
W_S	P_{VL} T_{VL} W_S	P_L T_{VL} W_S	P_H T_{VL} W_S	P_{VH} T_{VL} W_S
W_M	P_{VL} T_{VL} W_M	P_L T_{VL} W_M	P_H T_{VL} W_M	$P_{VH}T_{VL}$ W_M
W_F	P_{VL} T_{VL} W_F	P_L T_{VL} W_F	P_H T_{VL} W_F	$P_{VH}T_{VL}$ W_F

Table 10: Quartet Factors Array

	P_{VL} T_{VL} H_D	P_L T_{VL} H_D	P_H T_{VL} H_D	P_{VH} T_{VL} H_D
C_c	P_{VL} T_{VL} H_D C_C	P_L T_{VL} H_D C_C	P_H T_{VL} H_D C_C	P_{VH} T_{VL} H_D C_C
C_P	P_{VL} T_{VL} H_D C_P	P_L T_{VL} H_D C_P	P_H T_{VL} H_D C_P	P_{VH} T_{VL} H_D C_P
C_{CL}	P_{VL} T_{VL} H_D C_{CL}	P_L T_{VL} H_D C_{CL}	P_H T_{VL} H_D C_{CL}	P_{VH} T_{VL} H_D C_{CL}
W_S	P_{VL} T_{VL} H_D W_S	P_L T_{VL} H_D W_S	P_H T_{VL} H_D W_S	P_{VH} T_{VL} H_D W_S
W_M	P_{VL} T_{VL} H_D W_M	P_L T_{VL} H_D W_M	P_H T_{VL} H_D W_M	P_{VH} T_{VL} H_D W_M
W_F	P_{VL} T_{VL} H_D W_F	P_L T_{VL} H_D W_F	P_H T_{VL} H_D W_F	P_{VH} T_{VL} H_D W_F

The resulted factors from step 1 to step 4 represent the best rules that describe the hidden relations between the generated power and weather factors.

Step 4: Quintet factors array is generated. Each quartet factor levels generated from Quartet factors array is combined with the other factor level, therefore only one quintet factor levels is created for each element in the array. As an example, power.

Factor level such as (P_H T_{VH} H_D C_P) is combined with all other factors levels such as (W_F) to produce (P_H T_{VH} H_D C_P W_F) and not commutative process, so the trio factor levels (W_F P_H T_{VH} H_D C_P) are not allowed as shown in Table 11.

Table 11. Quintet Factors Array

	P_{VL} T_{VL} H_D C_C	P_L T_{VL} H_D C_C	P_H T_{VL} H_D C_C	P_{VH} T_{VL} H_D C_C
W_S	P_{VL} T_{VL} H_D C_C W_S	P_L T_{VL} H_D C_C W_S	P_H T_{VL} H_D C_C W_S	P_{VH} T_{VL} H_D C_C W_S
W_M	P_{VL} T_{VL} H_D C_C W_M	P_L T_{VL} H_D C_C W_M	P_H T_{VL} H_D C_C W_M	P_{VH} T_{VL} H_D C_C W_M
W_F	P_{VL} T_{VL} H_D C_C W_F	P_L T_{VL} H_D C_C W_F	P_H T_{VL} H_D C_C W_F	P_{VH} T_{VL} H_D C_C W_F

4. EXPERIMENTS AND RESULTS

The following measuring parameters are used in this paper to decide the best resulted rules:

a) Lift: P (A, B)/(P (A) P (B)) where Lift = 1 means A and B are independent. The larger the number (> 1), the more it indicates that the existence of A and B in a shopping basket is not accidental and has a strong correlation.

b) Leverage: P (A, B) -P (A) P (B) When Leverage = 0, A and B are independent, the greater the Leverage, the closer the relationship between A and B

c) Conviction: P (A) P (! B)/P (A, B) (! B means B did not happen). Conviction is also used to measure the independence of A and B. It can be seen from the relationship between it and lift (negating B and substituting Lift formula to find the reciprocal), the larger this value is, the more related A and B are.

A sample of the processed data to build the association rule model is shown in Table 12. The data was divided into three parts: training, testing, and validation. In the training part, 700 records are used to build the rule-based. After building the training model, 150 of the data is used to test the model, and the same parameter confidence is used for the most powerful measure of each rule.

Table 12: Processed Data To Build The Model

Temperature	Wind Speed	Humidity	Cloud Cover	Power
Very High	Slow	Comfort	Clear	very low
Very High	Slow	Comfort	Clear	very low
Very High	Slow	Comfort	Clear	very low
Very High	Slow	Comfort	Clear	very low
Very High	Slow	Dry	Clear	very low
Very High	Slow	Comfort	Clear	very low
Very High	Medium	Comfort	Clear	very low
Very High	Slow	Comfort	Clear	very low
Very High	Slow	Comfort	Clear	very low
Very High	Slow	Comfort	Clear	very low
Very High	Slow	Comfort	Clear	very low
Very High	Medium	Comfort	Clear	very low
Very High	Slow	Humid	Clear	very low
Very High	Medium	Humid	Clear	very low
High	Medium	Humid	Clear	very low
Very High	Slow	Dry	Clear	very low

The best achieved rules are shown in Table 13. The results are retrieved after setting the support to be 0.1 and confidence to be 0.7. Weaka [28] software 3.95 is used to build the proposed system. Weka software is developed by Waikato Environment for Knowledge Analysis.

It is developed at the University of Waikato, New Zealand. Weaka is an open-source software which provides tools for data preprocessing, implementation of several machine learning algorithms, and visualization tools. From the table it can be seen that the minimum confidence and lift are

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

0.7, 0.97 respectively.

Table 13: The Best Achieved Rules

Association Rules:
1. *If Power is very high and Temperature is very high, then Cloud Cover is Clear* (conf =1, lift = 1.33).
2. *If Power is very high and Humidity is Comfort and Wind is Slow, then Cloud Cover is Clear* (conf = 0.94, lift = 1.25).
3. *If Power is medium and Temperature is Low, then Wind is Slow (conf =0.87, lift = 1.17).*
4. *If Power is very low and Humidity is Humid, then Temperature is Low (conf = 0.85, lift = 1.97).*
5. *If Power is high and Humidity is Humid, then Wind is Slow (conf = 0.85, lift = 1.14).*
6. *If Power is high and Humidity is Humid, then Cloud Cover is Clear and Wind is Slow (conf = 0.75, lift = 1.3)*
7. If Power is very high and Humidity is Comfort, then Cloud Cover is Clear and Wind is Slow (conf = 0.72, lift = 1.25)
8. *If Power is very high and Temperature is High, then Wind is Slow (conf = 0.72, lift = 0.97).*
9. *If Power is very high, then Humidity is Comfort (conf = 0.72, lift = 1.41).*
10. *If Power is very high, then Cloud Cover is Clear and Wind is Slow (conf = 0.71, lift = 1.22).*
11. *If Power is very low and Temperature is Low, then Humidity is Humid (conf = 0.7, lift = 1.68).*
12. *If Power is very low, then Humidity is Humid (conf = 0.7, lift = 1.67).*

After generating the rule, a new 150 cases have not previously used in training and testing are used to validate the model. Two domain experts reviewed the results of weather decision model one by one. The resulted decision matches the generated decision 97.33%. This methodology can help the university and other similar sites, that cannot provide weather measuring devices, to use the generated power values from the photovoltaic system in predicting weather values.

5. CONCLUSION

In this paper, a new model using association rules mining algorithm has been developed to predict the weather factors. The developed model has a new strategy in predicting weather factors without measuring devices. The developed system can be considered as virtual weather station. In this model a real power PV data is used to predict the weather factors. The resulted rules which were extracted from the model are promising and show high confidence greater than 0.7. The system has been validated using 150 new cases. The achieved results match 97.33% of the domain expert results.

REFERENCES:

[1] Mansour, A.M., Abdallah, J., Obeidat, M.A.," An efficient intelligent power detection method for photovoltaic system," International Journal of Circuits, Systems and Signal Processing, vol. 14, pp. 686–699, 2020.

[2] Ayman M. Mansour, Murad M. Alaqtash, Mohammad Obeidat, "Intelligent Classifiers of EEG Signals for Epilepsy Detection," WSEAS Transactions on Signal Processing, vol. 15, 2019.

[3] Murad Alaqtash, Ayman M Mansour, Mohammad Obeidat, "Fuzzy Assessment Model for Functional Impairments in Human Locomotion". IOSR-JECE, vol. 14, no. 1, Jan-Feb 2019.

[4] Ayman M. Mansour, "Intelligent E-Health System for Patient and Elderly People Monitoring Using Multi Agents System," Jordan Journal of Electrical Engineering, vol. 4, no. 1, 2018.

[5] Ayman M. Mansour, "Decision Tree-Based Expert System for Adverse Drug Reaction Detection using Fuzzy Logic and Genetic Algorithm," International Journal of Advanced Computer Research (IJACR), vol. 8, no. 36, 2018.

[6] Mohammad A. Obeidat and Ayman M. Mansour, "EEG Based Epilepsy Diagnosis System using Reconstruction Phase Space and Naïve Bayes Classifier," WSEAS Transactions on Circuits and Systems, vol. 17, 2018.

[7] Mansour, A.M., Obaidat, M.A. and Hawashin, B. Elderly people health monitoring system using fuzzy rule based approach. International Journal of Advanced Computer Research, vol. 4, no. 4, p.904. 2014.

[8] Ayman M. Mansour, "GSM based Vehicle-to-Vehicle Communication using Multi-Agent Intelligent System," WSEAS Transactions on Electronics, vol. 10, 2019.

[9] A. M. Mansour and M. A. Obeidat, "Impact of Infertility Factors on IVF Success Rate using

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Fuzzy Inference System," 2021 International Conference on Information Technology (ICIT), 2021, pp. 914-918, 2021.

[10] Mansour, Ayman M., Abdulaziz Almutairi, Saeed Alyami, Mohammad A. Obeidat, Dhafer Almkahles, and Jagabar Sathik. "A Unique Unified Wind Speed Approach to Decision-Making for Dispersed Locations" Sustainability 13, no. 16: 9340, 2021.

[11] A. M. Mansour, M. A. Obeidat and J. Abdallah, "A Novel Multi-agent Mechanism for Managing Electrical Power Smart Grids," 12th International Renewable Engineering Conference (IREC), 2021, pp. 1-6, 2021.

[12] Ayman M. Mansour, "Texture Classification using Naïve Bayes Classifier," International Journal of Computer Science and Network Security (IJCSNS), vol. 18, no. 1, January 2018.

[13] Ayman M Mansour, "Cooperative Multi-Agent Vehicle-to-Vehicle Wireless Network in a Noisy Environment," International Journal of Circuits, Systems and Signal Processing, vol. 15, 2021.

[14] Suhail Sharadqah, Ayman M Mansour, Mohammad A Obeidat, Ramiro Marbello and Soraya Mercedes Perez, "Nonlinear Rainfall Yearly Prediction based on Autoregressive Artificial Neural Networks Model in Central Jordan using Data Records: 1938-2018" International Journal of Advanced Computer Science and Applications(IJACSA), 12(2), 2021.

[15] Agrawal, R., Imielinski, T. and Swami, A. Mining association rules between sets of items in large databases. In Proceedings of the ACM SIGMOD International Conference on Management of Data, 207-216, 1993.

[16] Qiankun Zhao and Sourav S. Bhowmick. Association rule mining: A Survey. Technical Report, CAIS, Nanyang Technological University, Singapore, No. 2003116, 2003.

[17] Agrawal, R. and Srikant, R. Fast Algorithms for Mining Association rules. Proc. 20th VLDB conference, Santiago, Chile, 1994.

[18] Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar, "Introduction to Data Mining," 2 nd Edition, Addison Wesley, Boston, MA, ISBN 978-0133128901, 2018.

[19] Han, J., Pei, J. and Yin, Y. Mining frequent patterns without candidate generation. ACM SIGMOD Intl. Conference on Management of Data, ACM Press, 1-12, 2000.

[20] Amitabha Das, Wee Keong Ng and Yew KWong Woon. Rapid Association Rule Mining. ACM, Pages 474–4811, 2001.

[21] Jong Soo Park, Ming-Syan Chen and P. S. Yu, "Using a hash-based method with transaction trimming for mining association rules," in IEEE Transactions on Knowledge and Data Engineering, vol. 9, no. 5, pp. 813-825, Sept.-Oct. 1997.

[22] P. Yazgana and A. Kusakci, "A Literature Survey on Association Rule Mining Algorithms," Southeast Europe Journal of Soft Computing, 5(1), 2016.

[23] C. Basu, M. Padmanaban, S. Guillon, L. Cauchon, M. DeMontigny and I. Kamwa, "Association rule mining to understand GMDs and their effects on power systems," 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, pp. 1-5, 2016.

[24] P. Xu, F. Xiao, S. Feng, Y. Wang and K. Ye, "Data mining of power transmission line fault based on apriori algorithm," 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, pp. 49-54, 2017.

[25] M. Cao and C. Guo, "Research on the Improvement of Association Rule Algorithm for Power Monitoring Data Mining," 10th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, pp. 112-115, 2017.

[26] M. Tian et al., "Data Dependence Analysis for Defects Data of Relay Protection Devices Based on Apriori Algorithm," in IEEE Access, vol. 8, pp. 120647-120653, 2020.

[27] Jiawei Han and Micheline Kamber. Data Mining: Concepts and Techniques (2nd. ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA,2006.

[28] Witten, I.H., Frank, E., Trigg, L., Hall, M., Holmes, G. & Cunningham, S.J, Weka: Practical machine learning tools and techniques with Java implementations. (Working paper 99/11). Hamilton, New Zealand: University of Waikato, Department of Computer Science, 1999.