IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)
e-ISSN: 2278-1676,p-ISSN: 2320-3331, Volume 13, Issue 5 Ver. II (Sep. – Oct. 2018), PP 39-48
www.iosrjournals.org

Grid Monitoring using Solar SCADA Dataset

[1]Syed FajarHazrat, [2]Pooja Khatri, [3]Muheet Ahmed Butt, [4]Majid Zaman
[1]Student, Department of Electrical Engineering, Swami Devi Dayal Institute of Engineering and Technology, Haryana, India
[2]Assistant Professor, Department of Electrical Engineering, Swami Devi Dayal Institute of Engineering and Technology, Haryana, India.
[3]Scientist "D",PG Department of Computer Sciences, University of Kashmir, Srinagar, J&K.
[4]Scientist "D", Directorate of IT and SS, University of Kashmir, Srinagar, J&K.
Corresponding Author:Syed FajarHazrat

Abstract: *Grid is framed as a planned nationwide network that makes use of information so as to deliver electricity in an efficient, reliable and secure manner. This entire electrical system is very difficult to monitor because any rise or fall of electrical impulses, in this system may result in the damage of highly expensive electrical components, as a result causes failure to an entire system. Therefore, there is dire need of controlling such damage of these synchronous machines and other component appliances so that these failures are avoided, which may result in loss of life and property. We have used various data mining approaches on a SCADA dataset which has been derived from solar electric power plant and is being monitored by grid station. The data pertains to the real time values pertaining to voltage and current. The proposed research uses this real time data for constructing a probabilistic model to respond to the above problem in a very intelligent manner.*

Keywords:*SCADA system, Naïve Bayes algorithm*

Date of Submission: 17-11-2018 Date of acceptance: 04-12-2018

I. Introduction

Grids using computer-based methodologies, information technology, communication technology etc are becoming important nowadays. Electrical power system is a network of electrical components, which is used to effectively supply and transfer electrical power. The grid station is an integral part of this electrical power system which allows us to transfer this power to an extended geographical area.

A grid-connected photovoltaic power system or grid-connected PV power system is an electricity-generating solar power system that is connected with a utility grid. A grid-connected PV system consists of the solar panels, inverters, a power conditioning unit and various grid-connecting equipment. Solar energy collected by photovoltaic solar panels, is conditioned or processed for use, by using grid-connected inverters after being delivered to a power grid. The inverter used in the system changes the DC input voltage from the PV to AC voltage for grid. Monitoring of grid voltage, waveform and frequency plays an important role. The inverter plays an important role in monitoring by not allowing solar energy in case the grid is dead or nominal specifications change. It also helps in the synchronization with the grid waveform, thereby producing voltage slightly higher than the grid itself, in order for smooth outward flow of energy from the solar array.

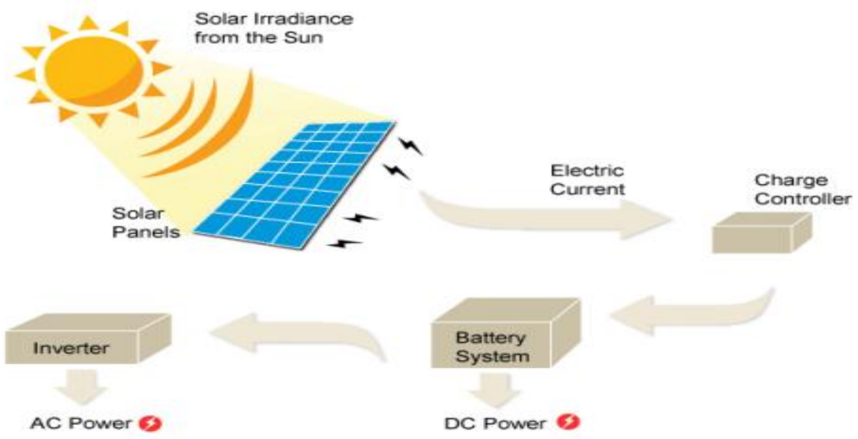

Fig. 1. Block diagram of electrical generation from sunlight.

The proposed smart system is to be placed between the input supply lines and the grid station. This smart system will sense the input power from various sources and will take a decision in anticipation, whether the power needs to be transmitted in grid or not. In this regard, an above mentioned learning set has been created on the basis of range of voltage and currents in an electrical power system. Naïve Bayes method probabilistic model has been used for constructing classifier. A supervised learning approach has been adopted and the parameter estimation for the Naïve Bayes has been done by using maximum likelihood method. The training data has been collected from local SCADA, which is around one lakh forty five thousand records. It has been observed when the system was placed in an appropriate order of 98.6% efficiency was calculated. Thus, predicting any abnormality arising in the system. Other data mining approaches were also implemented on the same data set and it was observed, that Naïve Bayes still gave better results.

Grid tripping

An electrical grid is an inter-connected network for delivering electricity from suppliers to consumers. Many of the grids are heavily-loaded and operating close to the maximum capacity. Due to sudden bulk transfer, the grid becomes unstable, resulting in tripping or blackout. Tripping happens in an unbalanced condition or it is the isolation of the part of the system to prevent damage to equipment. It may also occur during frequency drop or other unbalance condition. In case the demand increases or is more than power generated, it leads to the frequency drop. When this frequency becomes less than the certain limit, the grid will trip. This will further result in loss of power. Similar situations will result when the demand is less than generation.

Factors resulting in tripping:

There are various factors which influence the stability of system:

Voltage (V) :A voltage spike would typically not cause the AC line current in any appliance to increase, though it might potentially trigger a catastrophic failure within the appliance which results in uncontrolled current flow. Sometimes transient suppressors on the mains entry control panel are used to suppress the spikes. Also in case of voltage drop caused by excessive current draw through the circuit breaker will cause the breaker to trip, not because of the lower voltage but because of a higher current drawn than the ampere rating of the circuit breaker. A low voltage from the voltage source should not cause the circuit breaker to trip as long as the current draw of the load remains within the breaker's rating.

Current (I):Current is an important factor which is responsible for the tripping of circuit. The majority of circuit breakers are not voltage sensitive, they are current and current/time sensitive. Circuit breakers are becoming more accurate and have electronic current sensing in the most advanced units. Breakers look at two kinds of overcurrent. Magnetic trip, where a surge of current is significantly above circuit ratings. The circuit trips instantly. Thermal trip, where a mild overcurrent is threatening to eventually overheat the wires. The breaker allows this current for short term, but trips the circuit before wires may overheat which may cause fire. This will trip in several seconds to dozens of minutes, depending on the overload and according to the breakers trip curve.

Neutral current (I_n) :When neutral current is high, the star junction point of transformer from zero(ground) potential will shift due to considerable voltage drop, and as a result the voltage on heavily-loaded phase would reduce and low loaded phase would increase and may cause damage to appliances. If the neutral current is too high the connection may get burned and becomes open, giving rise to heavy voltage difference than normal voltage, causing black out and damage to appliances.

Short Circuit Current (I_{sc}): When the impedance is low, short circuit current I_{sc} corresponds to the short circuit condition and is calculated when the voltage equals zero. At $V = 0, I = I_{sc}$, which occurs at the beginning of the forward-bias sweep and is the maximum current value in the power quadrant. For forward-bias power quadrant, $I_{sc} = I_{max}$

Open Circuit Voltage (V_{oc}):When there is no current passing through the system, the open circuit voltage (V_{oc}) is occurred. At $I = 0, V = V_{oc}$, which is the maximum voltage difference across the circuit. For forward-bias power quadrant, $V_{oc} = V_{max}$.

Power Factor Correction: Power Factor, the ratio between the real power and the apparent power forms a very essential parameter in power system. It is indicative of how effectively the real power of the system has been utilized.

Active Power:Power is a measure of energy per unit time. Power therefore gives the rate of energy consumption or production. The units for power are generally watts (W). The total amount of energy consumed by this appliance is the wattage multiplied by the amount of time during which it was used, this energy can be expressed in units of watt-hours (or, more commonly, kilowatt-hours). The power dissipated by a circuit element whether an appliance or simply a wire is given by the product of its resistance and the square of the current through it$P = I^2R$. The term "dissipated" indicates that the electric energy is being converted to heat. Another, more general way of calculating power is as the product of current and voltage:$P = VI$. For a resistive element, we can apply Ohm's law ($P = VI$) to see that the formulas $P = I^2R$ and $P = VI$ amount to the same thing

$$P = VI = (IR)I = I^2R \quad (1)$$

Complex Power: Applying the simple formula $P = VI$ becomes more problematic when voltage and current are changing over time, as they do in A.C systems. In the most concise but abstract notation, power, current, and voltage are all complex quantities, and the equation for power becomes $S = VI^*$ where S is the apparent power and the asterisk denotes the complex conjugate of the current I, meaning that for purposes of calculation, the sign (positive or negative) of its imaginary component is to be reversed.. Regardless of all the complexities to be encountered, it is always true that the instantaneous power is equal to the instantaneous product of current and voltage. In other words, at any instant, the power equals the voltage times the current at that instant. This is expressed by writing each variable as a function of time, $P(t) = V(t) \times I(t)$where t is the same throughout the equation (i.e., the same instant).

Reactive Power: Reactive power is the component of power that oscillates back and forth through the lines, being exchanged between electric and magnetic fields and not getting dissipated. It is denoted by the symbol Q, and its magnitude is given by

$$Q = V_{rms} * I_{rms} * sin\phi \quad (2)$$

For the resistive case where $\phi = 0$ and $sin\phi = 0$as there will be no reactive power at all. Reactive power is measured in VAR (also written Var or VAr), for volt-ampere reactive.

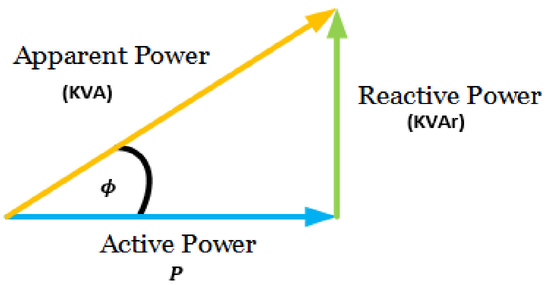

Fig. 2. Power triangle.

Power Factor: Power factor is a figure of merit that measures how effectively power is transmitted between a source and load network.

$$Powerfactor = \frac{(AveragePower)}{(RMSVoltage)(RMSCurrent)} \quad (3)$$

It always has a value between zero and one. The unity power factor condition occurs for a load that obeys Ohm's Law. In this ideal case, the voltage and current waveforms have the same shape, contain the same harmonic spectrum, and are in phase. The rms current and voltage can be minimized for a given average power throughput at maximum (unity) pf, *i.e.* with linear resistive load.

In non-ideal cases where the voltage waveform doesn't contain any harmonics and the load is nonlinear, the power factor is product of two terms, one arising from phase shift of the fundamental component of the current and other resulting from the current harmonics. The first term named as displacement factor and other term as distortion factor. In electrical concern, power factor is defined as the ratio of the active power P to the apparent power

$$pf = \frac{P}{S} \quad (4)$$

For purely sinusoidal voltage and current, the ideal definition is applied as,

$$pf = cos\phi \quad (5)$$

where$cos\phi$ is the displacement factor of the voltage and current.

II. Literature Survey

MAO Anjia, YU Jiaxi, GUO Zhizhong, "PMU placement and data processing in WAMS that complements SCADA", 2005 presented a paper regarding PMU placement method at the rule of observability of kernel network, which was simplified by the load zone classification. The system's operation states could be obtained by using fewer PMU's, by the use of PMU measurement information extended method. Also for the improvement of the precision of the processed data, state estimation was applied. This method made WMAS accurate data complement SCADA nom accurate data effectively, and led the WAMS and SCADA complement each other in practice [1].

T.Sanislav, D. Capatina, L. Miclea, "A data mining experiment on a SCADA system's historical acquired data, 2008 presented a paper which gave the implementation for a hydroelectric power plants cascade, monitoring, controlling and over-limit alarm processes in a SCADA system which provided large amount of historical data which could be stored in distributed databases. This data could not offer a performance knowledge discovery in the preliminary form. However in order to extract hidden information directly from SCADA system's database, intelligent data mining framework could be applied over historical acquired data. Thus, this paper presented a data mining and visualization experiment that could be performed on a real data and therefore, results were achieved in the experiment [2].

Richa Gupta, Moinuddin, Parmod Kumar, "Cloud computing data mining to SCADA for energy management", 2015 presented a paper regarding cloud mining infrastructure based on SAP HANA for SCADA. It gave a powerful decision taken tool that could be operated at energy control center. Important features regarding hardware and software of HANA were also defined. It also gave a configuration of SCADA for energy management system, which was based on cloud computing mining. It also described cloud data mining features and service model of cloud [3].

NursedaYildirim and BahriUzunoglu, "Association rules for clustering algorithms for data mining of temporal power ramp balance", 2015 presented a paper in which k-means clustering and association rules of Apriori algorithm were implemented for the analyzation and prediction of wind power ramp occurrences which was based on 10 minutes temporal SCADA data of power from records of Ayyildiz wind farm. Power ramps could also be computed from this data. Based on temporal data five wind turbines with no dissimilarity measure in space was clustered. The power ramp data was analyzed by the K-means algorithm for the calculation of their cluster means and cluster labels. Association rules of data mining algorithm were employed for the analyzation of temporal ramp occurrences between wind turbines [4].

W.Alves¸ D. Martins, U. Bezerra and A. Klautau, "A hybrid approach for big data outlier detection from electric power SCADA system", 2017 presented a paper which aimed at circumventing these restrictions by giving a methodology that dealed with SCADA big data that consisted of a pre- processing algorithm and hybrid approach outlier detectors. The hybrid approach was assessed using real data that was taken from Brazilian utility company. The results showed that the proposed methodology was capable of identifying outliers that correlated with important events that affected the system [5].

RamazanBayindir, MehnetYesilbudak, MedineColak, Naci Gene, "A novel application of Naïve Bayes classifier in photovoltaic energy prediction", 2017 has presented a paper which predicted the daily total energy generation of an installed photovoltaic system using the Naïve Bayes classifier. In this prediction process, one year historical dataset which included daily average temperature, daily total sunshine duration, daily total global solar radiation and daily total photovoltaic energy generation parameters were used as a categorical valued attributes. The sensitivity and accuracy improved for the photovoltaic energy prediction and the effects of other solar attributes on the photovoltaic energy generation were evaluated using Naïve Bayes application [6].

Lei Chen, Wei Lu, Liqiangwang, ErgudeBao, Weiwei Xing, Yong Yang, Victor Yuan, "Optimizing map reduce partitioner using Naïve Bayes classifier", 2017, presented a paper which proposed a novel partitioner which was based on Naïve Bayes classifier, namely, BAPM, which achieved better performance through optimization of data locality and data skewing by leveragation of the Naïve Bayes classifier, i.e., consideration of job type and bandwidth as classification attributes [7].

Sadia Ansari, Kanchan Hans, Sunil Kumar Khatri, "A Naïve Bayes classifier approach for detecting hypervisor attacks in virtual machines", 2017 presented a paper which proposed a framework to detect the hypervisor attacks in virtual machines. Bayesian classifier on the publicly available dataset was used. Characterization vulnerabilities of two hypervisors XEN and VMware, based on real-time attacks was done. Three attribute namely authentication, integrity impact and confidentially impact were considered for the input feature vector. Calculation of posterior probability of a vulnerability which indicated the degree of it being a hypervisor attack was done [8].

Sebastian Romy Gomes, SK Golam Saroar, MD MosfaiulAlamTelot, BehrozNewaz Khan, "A comparative approach to email classification using Naïve Bayes classifier and hidden Markov Model", 2017 presented a paper which investigated a comparison between two different approaches for classification of emails based on their categories. Naïve Bayes classifier and Hidden Markov Model (HMM), two different machine learning algorithms, both had been used for the detection that whether the email is important or spam. Various combinations of NLP techniques-stop words removing, stemming, lemmatizing had also been tried on both the algorithms for the inspection of the differences in accuracy as well as for finding the best method among them [9].

Basel Alshaikhdeeb and KamsuriahAhmad , "Feature selection for chemical compound extraction using Wrapper Approach with Naive Bayes Classifier", 2017 gave a paper which aimed to apply a combination of Naïve Bayes classification method with the Wrapper Subset Selection approach so as to identify the best features. Results obtained showed that the proposed combination had the ability for the identification of the best combination of features which consisted of capitalization, punctuation, prefix and part-of-speech tagged by achieving 0.72 of f-measure. Such result had been compared to the state of the art and it demonstrated competitive performance [10].

Ali HaghpanahJahromi and Mohammad Taheri, “A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features”, 2017 gave a paper which presented a new group of Gaussian naive Bayes classifiers which was based on the mixture of Gaussian distributions which was formed on less conditional dependent features that was extracted by local PCA. A semi-Ada Boost approach is used for dynamic adaptation of distributions which considered misclassified instances. Evaluation of proposed method had been done and compared with the related work on 12 UCI machine learning datasets and achievements showed significant improvement on the performance [11].

Peixin Liu, Hongzhi Yu, Tao Xu, Chuanqi Lan, “Research on archives text classification based on Naïve Bayes”, 2017 gave a paper which analyzed the data resources of archives in Gansu Province by combination with the characteristics of archives resources and combined with Naïve Bayesian classification algorithm so as to realize the application of archives resource classification [12].

Naive Bayes Approach

In machine learning, Naive Bayes classifiers are a family of simple "probabilistic classifiers" based on the application of Bayes' theorem with strong (naive) independence assumptions between the features. Naive Bayes has been studied extensively since the 1950s. With appropriate pre-processing, it is competitive in different domains with more advanced methods including support vector machines. It also finds its application in automatic medical diagnosis. In the statistics and computer science literature, Naive Bayes models are known under a variety of names, including simple Bayes and independence Bayes. Thus, is a simple technique for constructing classifiers, models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle, “all Naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable.” Bayes’ theorem provides a way that we can calculate the probability of a hypothesis given our prior knowledge.Bayes’ Theorem is stated as:

$$P(h|d) = (P(d|h) * P(h)) / P(d) \tag{6}$$

Where,

$P(h|d)$ is the probability of hypothesis h given the data d. This is called the posterior probability.

$P(d|h)$ is the probability of data d given that the hypothesis h was true.

$P(h)$ is the probability of hypothesis h being true (regardless of the data). This is called the prior probability of h.

$P(d)$ is the probability of the data (regardless of the hypothesis).

After calculating the posterior probability for a number of different hypotheses, we can select the hypothesis with the highest probability. This is the maximum probable hypothesis and may formally be called the maximum a posteriori (MAP) hypothesis.

This can be written as:

$$MAP(h) = max(P(h|d)) \tag{7}$$

or

$$MAP(h) = max((P(d|h) * P(h)) / P(d)) \tag{8}$$

or

$$MAP(h) = max(P(d|h) * P(h)) \tag{9}$$

The P(d) is a normalizing term which allows us to calculate the probability.

General Methodology of Proposed Research

The basis of learning set is based on the range of voltages and currents taken. Four set of values have been assigned for three phases of voltages as low, medium, normal and abnormal. Similarly in case of three phases of currents, we have minimum, normal and maximum range assigned. The below mentioned tables 1, 2, 3 and 4 provides a rule based structure on which the main dataset and the learning set is to be evaluated. This primary dataset is the preliminary dataset for generating the actual SCADA for the monitoring and control system.

Table 1. Voltages values for VrKv and VyKv

Grid name			VR KV					VyKV	
	VR kv low	**Vr KV MED**	**Vr KV normal**	**Vr KV ABNOR**		**Vy KV Low**	**Vy KV Med**	**Vy KV normal**	**VyKV ABNOR**
132 kvChesmashi 1st	51.1	66.26	76.2	92.2		51.1	65.68	76.2	92.2
132kv Chesmashi 2nd	51.1	65.44	76.2	92.2		51.1	65.57	76.2	92.2
132Kv ICT Pampore 1st	51.1	65.06	76.2	92.2		51.1	64.06	76.2	92.2
132Kv Pampore 2nd	51.1	63.02	76.2	92.2		51.1	65.28	76.2	92.2
132 Kv side Transformer bank 1st	51.1	66.14	76.2	92.2		51.1	64.59	76.2	92.2
132 Kv side Transformer bank 2nd	51.1	66.1	76.2	92.2		51.1	64.58	76.2	92.2
132 Kv side Transformer bank 3rd	51.1	65.05	76.2	92.2		51.1	64.57	76.2	92.2
220kv Mir bazar 1st	92.99	119.13	127.01	153.67		92.99	120.64	127.01	153.67
220kv Mir bazar 1st	92.99	119.12	127.01	153.67		92.99	120.66	127.01	153.67
220 kv transformer bank 1st	92.99	119.1	127.01	153.67		92.99	120.63	127.01	153.67
220 kv transformer bank 2nd	92.99	120.64	127.01	153.67		92.99	119.71	127.01	153.67
220 kv transformer bank 3rd	92.99	119.16	127.01	153.67		92.99	120.23	127.01	153.67
220 kvWagora 1st	92.99	119.17	127.01	153.67		92.99	120.29	127.01	153.67
220 kvWagora 2nd	92.99	121.78	127.01	153.67		92.99	122.05	127.01	153.67

Table2. Voltage and Current values for VbKv and R-ph

Grid Name			VbKV				R-ph	
	Vb KV Low	**VbKv MED**	**Vb KV normal**	**VbKV ABNOR**		**R- ph min current**	**R-ph Line current**	**R-ph Maxcurrent**
132 kvChesmashi 1st	51.1	65.19	76.2	92.2		0	436.61	500
132kv Chesmashi 2nd	51.1	65.19	76.2	92.2		0	355.98	500
132Kv ICT Pampore 1st	51.1	65.07	76.2	92.2		0	511.96	500
132Kv Pampore 2nd	51.1	64.82	76.2	92.2		0	499.89	500
132 Kv side Transformer bank 1st	51.1	64.81	76.2	92.2		0	593.83	500
132 Kv side Transformer bank 2nd	51.1	64.82	76.2	92.2		0	590.59	500
132 Kv side Transformer bank 3rd	51.1	65.06	76.2	92.2		0	616.71	500
220kv Mir bazar 1st	92.99	119.82	127.01	153.67		0	69.3	500
220kv Mir bazar 1st	92.99	119.85	127.01	153.67		0	68.84	500
220 kv transformer bank 1st	92.99	119.82	127.01	153.67		0	338.87	500
220 kv transformer bank 2nd	92.99	119.26	127.01	153.67		0	338.83	500
220 kv transformer bank 3rd	92.99	119.88	127.01	153.67		0	352.6	500
220 kvWagora 1st	92.99	119.93	127.01	153.67		0	568.63	500
220 kvWagora 2nd	92.99	120.69	127.01	153.67		0	569.88	500

Table3. Current values for Yph and Bph

		Y-ph				B- ph	
	Y-ph	**Y-ph Line**	**Y-ph**		**B-ph**	**B- ph Line**	**B-ph**
	min current	**Current**	**max current**		**min current**	**current**	**max current**
132 kvChesmashi 1st	0	432.47	500		0	424.7	500
132kv Chesmashi 2nd	0	354.28	500		0	344	500
132Kv ICT Pampore 1st	0	510.75	500		0	507.95	500
132Kv Pampore 2nd	0	504.64	500		0	499.37	500
132 Kv side Transformer bank 1st	0	590.45	500		0	589.23	500
132 Kv side Transformer bank 2nd	0	586.71	500		0	581.45	500

132 Kv side Transformer bank 3rd	0	622.86	500		0	605.66	500
220kv Mir bazar 1st	0	63.45	500		0	79.79	500
220kv Mir bazar 1st	0	64.62	500		0	79.11	500
220 kv transformer bank 1st	0	336.69	500		0	335.5	500
220 kv transformer bank 2nd	0	335.72	500		0	333.31	500
220 kv transformer bank 3rd	0	355.1	500		0	346.77	500
220 kvWagora 1st	0	534.1	500		0	557.15	500
220 kvWagora 2nd	0	581.35	500		0	568.04	500

Table 4. Different values for neutral current

		Neutral current								
	Min	Normal	Max	abnormal		Active power(MW)	React power(MW)	App Power (MVA)	Avg PF	Frequecy
132 kvChesmashi 1st	0	2.96	10.47284	12.67213		-80.56	-27.19	85.02	0.947	50.054
132kv Chesmashi 2nd	0	6.18	11.22695	13.58461		-65.89	-20.53	69.01	0.954	50.057
132Kv ICT Pampore 1st	0	4.65	3.562597	4.310743		-95.28	-28.16	99.35	0.958	50.062
132Kv Pampore 2nd	0	11.29	5.030199	6.086541		-92.78	-27.63	96.8	0.958	50.052
132 Kv side Transformer bank 1st	0	9.69	4.127517	4.994296		-110.1	-35.29	115.61	0.952	50.058
132 Kv side Transformer bank 2nd	0	6.18	7.945489	9.614042		-109.06	-35.29	114.62	0.951	50.058
132 Kv side Transformer bank 3rd	0	6.54	15.09578	18.26589		-114.63	-34.62	119.73	0.957	50.023
220kv Mir bazar 1st	0	1.71	14.33977	17.35113		-23.94	8.1	25.38	0.946	50.061
220kv Mir bazar 1st	0	6.57	12.90816	15.61888		-24.12	8.1	25.38	-0.946	50.059
220 kv transformer bank 1st	0	5.64	2.960186	3.581825		109.85	51.17	121.17	0.906	50.057
220 kv transformer bank 2nd	0	4.51	4.793256	5.799839		109.87	50.25	120.81	0.909	50.061
220 kv transformer bank 3rd	0	1.6	7.403641	8.958406		115.52	51.05	126.28	0.914	50.054
220 kvWagora 1st	0	21.96	30.45828	36.85452		187.28	66.73	198.81	0.942	50.048
220 kvWagora 2nd	0	30.37	12.49205	15.11538		197.48	67.99	208.86	0.946	50.054

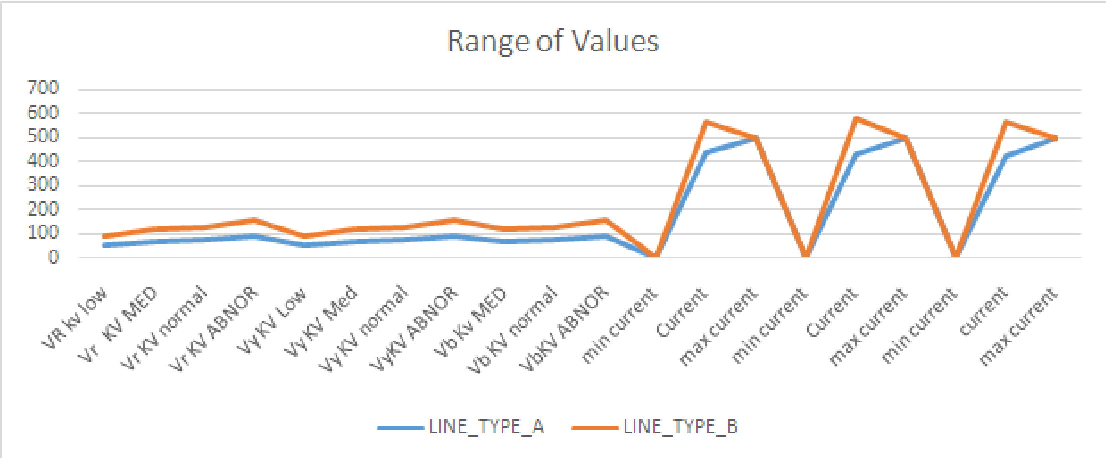

Fig. 3. Range of values for voltages and current

Attribute Information:

1. Vr KV: This is the line to line voltage, i.e, the voltage between two phases called as Line voltage.
2. Vy KV: This is the line to line voltage, i.e, the voltage between two phases called as Line voltage.
3. Vb KV: This is the line to line voltage, i.e, the voltage between two phases called as Line voltage.
4. R-ph: It is the line current through any one line between a three-phase source and load.
5. B-ph: It is the line current through any one line between a three-phase source and load
6. Y-ph: It is the line current through any one line between a three-phase source and load.

7. Neutral current: Under normal operating conditions, some phase unbalance occurs resulting in a small neutral current.

Main Dataset Organization

The data set was collected from the central grid station of Kashmir in an excel format. This format was migrated into a MS SQL server. The front analysis for the data set was carried out using dot framework. The main data set is composed of 145000 records.

Wagoora generating plant having capacity of 945/1260 MVA, 400/220kv, Mir Bazar 220/132 Kv, Pampore 220/132Kv and Chesmashahi grid 132/33Kv, are these four grids are connected in a ring circuit. These four grids which are connected to a smart monitoring and control unit, which enables the controlling of all these four grids, which in-turn are connected to other grids and sub-station. Incoming lines connected to the busbars. Each incoming lines is capable of supplying the grid station load. All these lines can be loaded simultaneously to share the grid station load or one line can be called upon to meet the entire load .This arrangement increases the reliability of the system. In case there is breakdown of incoming line, the continuity of supply can be maintained by other lines. The Grid Station has Double busbar system, one main busbar and the other reserve busbar. The incoming lines can be connected to either of the busbar with the help of arrangement of isolators and circuit breakers. The advantage of a double busbar system is that if repair is to be carried on one busbar, the supply need not to be interrupted as the entire load can be transferred to other bus.

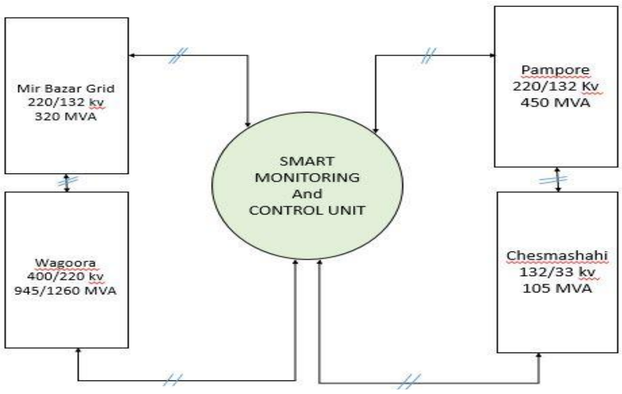

Fig. 4. Smart Monitoring and controlling unit

The main data set for the test is decomposed into two sub data sets. One is learning set and another is compare set. Out of 145000 values of voltages and currents, 135000 values have been made as learning set for Naïve Bayes and becomes the base for the any unknown value of voltages and currents that can be fed to the system. These remaining 10000 values of 145000 is to be fed to system for unknown monitoring and controlling results. The Naïve Bayes classifier generates the output status for these 10000 unknown values of voltage and current either working or tripped as per the situation. Thus, in this way we can predict any abnormality that might occur to the system, thereby protecting the entire system from any danger or blackout. The proposed monitoring system takes real time data from the incoming electric lines in the form of voltages and current. The system immediately compares these values with the training set values and forwards the relevant control signal to the grid. If any faulty data that may produce any abnormality leading to malfunctioning of the components, the system immediately takes suitable action so that these abnormalities are controlled.

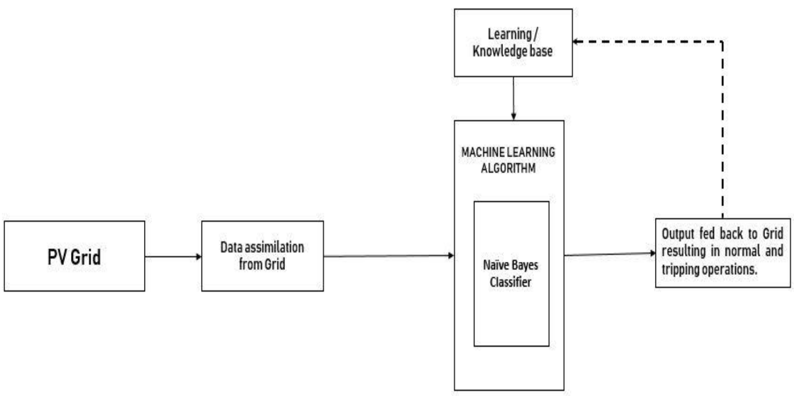

Fig. 5. Block diagram of proposed method

Proposed Algorithm

The algorithm is written using C # and has the following steps.

1. Algorithm for Data selection on collected data(this process is done only once):

a) Get real time grid data from SCADA sensor.

b) Migrate SCADA data into MS SQL Server (learning set data base), after proper validation based on generated values of voltages and current.

c) A correction code is also executed on the above learning set so as to remove outliers, insert missing values and also organize data inappropriate manner.

2. Algorithm for monitoring and controlling:

a) The data is streamed directly from the SCADA input.

b) This data is stored temporary in a table where the status weights are converted into numerical values and same are reflected in the table.

c) The said data is passed to a Naïve Bayes classifier where this data gets compared with an existing knowledge base.

d) In this process, the data is visualized for any abnormality which is reflected by triggering or non- triggering of grid.

e) The newly generated occurrence is again compared with the knowledge base and any deviation is recorded and saved.

III. Conclusion

A grid connected photovoltaic power system or grid connected PV power system is an electricity generating solar power system that is connected with a utility grid.The real time grid data is collected from SCADA sensor and the data is remotely or locally stored. This SCADA data stored on a local storage is migrated into MS SQL Server, after proper validation and correction codes on generated values of voltages and current.A migrated data is transformed into a learning set after performing necessary cleaning of outliers, missing values and also organized data inappropriate manner.This data is stored temporary in a database table where the status weights are converted into numerical values and same are reflected in the table.The said data is passed to a Naïve Bayes classifier where this data gets compared with an existing knowledge base/ Learning Set. In this process, the data is visualized for any abnormality which is reflected by triggering or non- triggering of grid. The newly generated occurrence is again compared with the knowledge base and any deviation is recorded and saved.Thus, in this way we can predict any abnormality that might occur to the system, thereby protecting the entire system from any danger or blackout.

In consequence of the overall prediction tests, the proposed Naive Bayes Approach for two way grid monitoring usingsolar SCADA datasetachieved the effective and efficient performance with the accuracy value of 98.99%.

References

[1]. MAO Anjia, YU Jiaxi, GUO Zhizhong. PMU placement and data processing in WAMS that complements SCADA, Power Engineering society general meeting, 2005.IEEE,780-783, 2005. http://ieeexplore.ieee.org.

[2]. T.Sanislav, D. Capatina, L. Miclea. A data mining experiment on a SCADA system's historical acquired data, Automation, Quality and Testing, Robotics, 2008. AQTR 2008.IEEE International Conference on 3, 419-422, 2008.http://ieeexplore.ieee.org.

[3]. Richa Gupta, Moinuddin, Parmod Kumar. Cloud computing data mining to SCADA for energy management, India Conference (INDICON),2015, Annual IEEE 1-6,2015.http://ieeexplore.ieee.org.

[4]. NursedaYildirim and BahriUzunoglu. Association rules for clustering algorithms for data mining of temporal power ramp balance, Cyberworlds (CW), 2015 International Conference on, 224-228, 2015.http://ieeexplore.ieee.org.

[5]. W.Alves¸ D. Martins, U. Bezerra and A. Klautau. A hybrid approach for big data outlier detection from electric power SCADA system, IEEE Latin America Transactions 15(1), 57-64, 2017.http://ieeexplore.ieee.org.

[6]. RamazanBayindir, MehnetYesilbudak, MedineColak, Naci Gene. A novel application of Naïve Bayes classifier in photovoltaic energy prediction, Machine Learning and Applications (ICMLA), 2017 16th IEEE International Conference on, 523-527, 2017. http://ieeexplore.ieee.org.

[7]. Lei Chen, Wei Lu, Liqiangwang, ErgudeBao, Weiwei Xing, Yong Yang, Victor Yuan. Optimizing map reduce partitioner using Naïve Bayes classifier, Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science, 2017.http://ieeexplore.ieee.org.

[8]. Sadia Ansari, Kanchan Hans, Sunil Kumar Khatri. A Naïve Bayes classifier approach for detecting hypervisor attacks in virtual machines, Telecommunication and Networks (TEL-NET), 2017 2nd International Conference on 1-6, 2017. http://ieeexplore.ieee.org.

[9]. Sebastian Romy Gomes, SK Golam Saroar, MD MosfaiulAlamTelot, BehrozNewaz Khan. A comparative approach to email classification using Naïve Bayes classifier and hidden Markov Model, Advances in Electrical Engineering (ICAEE), 2017 4th International Conference on, 482-487, 2017. http://ieeexplore.ieee.org.

[10]. Basel Alshaikhdeeb and Kamsuriah Ahmad. Feature selection for chemical compound extraction using Wrapper Approach with Naive Bayes Classifier, Electrical Engineering and Informatics (ICEEI), 2017 6th International Conference on, 1-6, 2017. http://ieeexplore.ieee.org.

[11]. Ali HaghpanahJahromi and Mohammad Taheri. A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features, Artificial Intelligence and Signal Processing conference (AISP), 2017, 209-212, 2017. http://ieeexplore.ieee.org.

[12]. Peixin Liu, Hongzhi Yu, Tao Xu, Chuanqi Lan. Research on archives text classification based on Naïve Bayes, Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2017 IEEE 2nd Information, 187-190, 2017.http://ieeexplore.ieee.org.
[13]. ApexaSuryakantPurohit, Chinmay Jani. Cascade tripping of the Power Grid, IJSRD International journal for Scientific Research and Development| Vol. 2, Issue 02, 2014. www.ijsrd.com
[14]. NPTEL lecture 51 on smart grid google.co.in.http://www.nptel.ac.in
[15]. Answers from www.quora.com.

IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) is UGC approved Journal with Sl. No. 4198, Journal no. 45125.

Syed FajarHazrat. "Grid Monitoring using Solar SCADA Dataset." IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) 13.5 (2018): 39-48.