ANALYSIS OF PERFORMANCE DATA FROM HVAC COMPONENTS FOR PREDICTION OF MAINTENANCE REQUIREMENTS

Ena Tobin, BEC/MSc, e.c.tobin@umail.ucc.ie
Hang Yin, BEng, hang.yin@umail.ucc.ie
Karsten Menzel, PhD, k.menzel@ucc.ie
University College Cork (UCC), Cork, Ireland

ABSTRACT

This paper describes a methodology which manages a building's maintenance activities by focusing on the timing of maintenance activities. Its goal is to optimise the trade-off between cost, which is incurred through maintenance activities, and the components health, which varies as a result of maintenance frequency. Here existing data from a BMS is utilises and analysis is performed on this data, with the objective of scheduling maintenance for a component, based on the measured performance of that component. This paper will investigate which data analysis technique provides the most certainty when determining the expected performance level. The major outcome of this paper is to present the certainty levels for each data analysis technique and illustrate how the analysis can be used for predicting maintenance requirements. Also this paper will have presented a methodology for managing maintenance activities and an implementation of these results using a Decision Support Framework for maintenance management. This research is performed as part of a nationally funded project 'Information Technology for Optimised Building Operation' (ITOBO).

Keywords: Energy-efficient buildings, Maintenance Management, Performance Based Maintenance, Performance Data Analysis, ITOBO.

1 INTRODUCTION

One of the main financial burdens of a building, during its life cycle, is the cost of maintenance. There are many different methods for performing maintenance activities, reactive, scheduled, condition-based maintenance. These methods make a trade-off between equipment health, cost and user-comfort. For example, with reactive maintenance, there can be extended down-time and loss of equipment health as the component may incur a complete failure before any maintenance activities are performed. On the other hand, for scheduled maintenance, unnecessary maintenance may be performed and so unnecessary costs and loss of manpower hours. In industrial applications, a more comprehensive form of maintenance is employed, such as condition-based maintenance. This is course is followed due to high costs associated with equipment down-time. For the most part, occupied office or educational facilities are not subjected to such stringent maintenance controls, but many of these buildings have the necessary data, already, to perform such maintenance activities. This data is provided from Building Management Systems (BMS) and can be used in conjunction with other tools to manage more effectively maintenance activities.

Firstly, in this paper, a generic methodology, which can be applied to any component, is described briefly. This can be found in more detail in (Tobin et al. 2010). It involves the specification of systems to be maintained; relationships, constraints and influences between these systems and components; and creation of a server to manage maintenance task information and descriptions. The methodology for determining the performance values for a component, at any point in time, will then be introduced. This is realised through utilisation of data warehousing technologies (DWH) and data mining techniques. A number of different algorithms are utilised to predict the performance of a component.

These algorithms will then be investigated and compared, with respect to which one provides the best results, in terms of predictive power. The Environmental Research Institute (ERI), on University College Cork campus, is utilised for an implementation of this methodology. This implementation will focus on predicting the performance of flat plate solar panels, which provide domestic hot water for the ERI, and also act as a back-up heat source for an under floor heating system. These results can then be used to directly schedule maintenance tasks or they can be used as an input to a Decision Making Framework (Yin et al. 2011), whereby evaluation of whether to perform maintenance tasks or to begin renovation activities can occur.

As specified in (EC 2010), we have an obligation to reduce our energy usage by 20% by 2020. As buildings consume 40% of global energy usage, one such way of achieving this goal is by ensuring our buildings work efficiently and without wasting energy. This means keep building service components in good working order and so a good maintenance management methodology is required.

Also a factor which motivates this research is that renovation is now becoming an attractive alternative to new construction. This means there is an even greater need for components to be maintained appropriately and for plans to be introduced to gauge when components should be replaced. This area will be discussed towards the end of this paper.

2 MAINTENANCE MANAGEMENT

2.1 Introduction to Maintenance Management

Maintenance management, as defined in (EN 13306 2001), includes all the activities of the management that determine:

- maintenance objectives or priorities (defined as targets assigned and accepted by the management and maintenance department)
- strategies (defined as a management method in order to achieve maintenance objectives),
- responsibilities and implement them by means such as maintenance planning, maintenance control and supervision, and several improving methods including economical aspects in the organization

Within maintenance management, it is necessary to choose a type of maintenance planning to utilize. There are a number of different types of maintenance which are used at present. They are (Tobin et al 2010):

1. Routine Maintenance, which is simple, small-scale, ongoing activities associated with regular and general upkeep of equipments, machines, plant, or system against normal wear and tear.
2. Emergency Maintenance is when urgent activities for sudden and unexpected failure of system or equipment, are performed.
3. Corrective Maintenance is when repair is done when a component has failed or broken down, to bring it back to working order.
4. Testing or failure-finding is finding out whether an item is able to work if required to do so on demand.
5. Predictive Maintenance is evaluation of the condition of equipment by performing periodic or continuous equipment condition monitoring.
6. Performance Based Maintenance is using sensors and meters to monitor the performance of equipment and schedule maintenance activities according to performance trends.

This last type of maintenance, PBM, is the type which we will utilise for this paper. PBM can be further classified as a meeting point between scheduled and reactive maintenance. It allows for faults which were not foreseen at the time of scheduling to be dealt with before they reach failure, i.e. before reactive maintenance is required. Cost allocation can be carried out for performance based but not for reactive maintenance.

2.2 Prognosis

The overall method to analyse these data coming from components and identify performance trends is for this research linked to prognosis. Here, the data classification methods, which are used in prognosis, are utilised.

Prognostics is detecting the precursors of a failure and predicting how much time remains before a likely failure (Gu et al. 2007). It is a method that enables monitoring the state of reliability of a product in real time, and therefore can be used to provide advance warning of a failure to minimise unscheduled maintenance, to provide condition-based maintenance and to help in product design and development (Schwabacher et al. 2007).

There are a number of different prognosis types, such as: model based, data driven and hybrid models. There are numerous algorithms the most widely used are: neural networks, anomaly detection algorithms/outlier detection algorithms, reinforcement learning, classification, clustering, Bayesian methods, fuzzy logic and Dempster-Shafer theory (Schwabacher et al. 2007).

Also there are a number of categories which prognosis can be divided into a number of categories, they are (Vichare et al. 2006; Gu et al. 2007):

- Using expendable prognostic cells that fail earlier than the host product to provide advance warning of a failure
- Monitoring and reasoning of parameters, such as shifts in performance parameters, progression of defects, that are precursors to impending failure
- Modelling stress and damage in electronics utilizing exposure conditions coupled with physics-of-failure (PoF) models to compute accumulated damage and assess remaining life.

3 DATA ANALYSIS FOR PERFORMANCE PREDICTION

In order to extract valuable information from BMS or wireless sensed data, it is necessary to clean, filter and analyse the data. To achieve this, the type and behaviour of the data must be known, the appropriate techniques must be used to acquire relevant and realistic information.

There are many methods available for analysing data, it is necessary to choose the type which is suitable for the analysis data and one that is also suitable for the expected results. A number of different techniques used are for this implementation, they are described by (Oracle 2005) as:

3.1 Logistical Regression (Generalised Linear Modelling – GLM)

This type of analysis includes a number of assumptions, such as, the target ‘is normally distributed conditioned on the value of predictors with a constant variance regardless of the predicted response value’.

Some of the advantages of this type of model are ’computational simplicity, an interpretable model form, and the ability to compute certain diagnostic information about the quality of the fit’. GLM is a parametric modeling technique.

3.2 Support Vector Machines

Support Vector Machines (SVM) ‘has strong theoretical foundations based on the Vapnik-Chervonenkis theory’. It also has ‘strong regularization properties’. ‘Regularization refers to the generalization of the model to new data’.

3.3 Naïve Bayes

The Naïve Bayes algorithm ‘is based on conditional probabilities’. Bayes' Theorem is the foundation of this technique. It uses a formula that ‘calculates a probability by counting the frequency of values and combinations of values in the historical data’. It is as follows:

Prob(B given A) = Prob(A and B)/Prob(A)

These four data analysis techniques will be implemented with the goal of predicting the performance of a component.

4 METHODLOGY – PERFORMANCE BASED MAINTENANCE

4.1 Overall Scenario for Managing Maintenance Activities

To manage maintenance appropriately and to ensure efficient use of energy, money, and man power, it is necessary to provide a holistic scenario for managing maintenance activities. This includes an overall system, which manages both maintenance personnel and equipment. This scenario has been presented in numerous papers, (Tobin et al. 2010b) and is funded by an SRC project titled 'Information Technology for Optimised Building Operation'. The scenario is as follows:

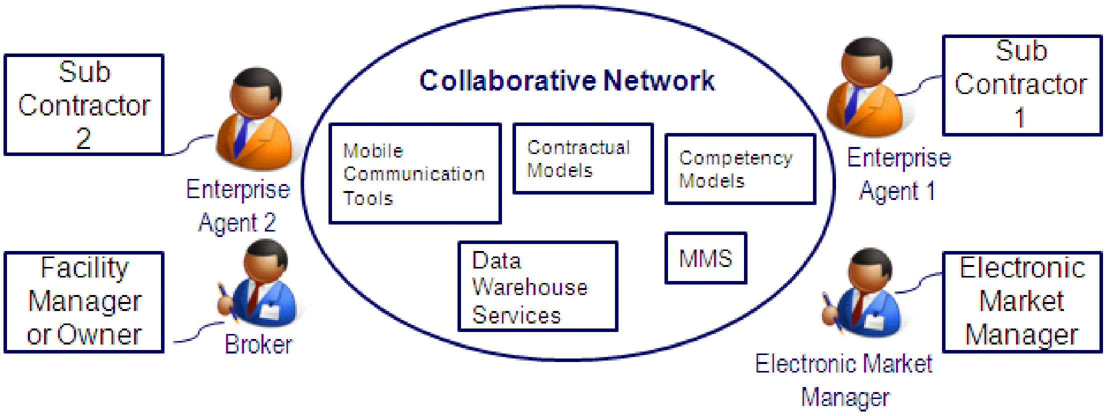

Figure 1 Overall maintenance management scenario

A **collaborative network** is required, including specification of data access rights and communication protocols. This specifies how actors or enterprises interact within the network and controls the flow of information.

A **server** is required to store all maintenance task descriptions. This is achieved through use of ARIS methodology and these diagrams are utilised to provide information to a scheduling tool, such as: what competencies are required to perform the maintenance; what equipment or resources are required; and what constraints are related to each task.

A system is required for managing all facility management activities (a **Maintenance Management System**). This system will manage the interactions through using the specifications of the collaborative network and will provide a link between the scheduler, server and Data Warehouse (DWH).

4.2 Performance Based Maintenance Scheduling Scenario

This research paper focuses now on the actual scheduling activities, at the appropriate time, and how this can be integrated into the overall scenario. Here, a methodology is presented on how to achieve this and on how to evaluate the different techniques which can be used. There are a number of steps in order to schedule activities based on a components performance:

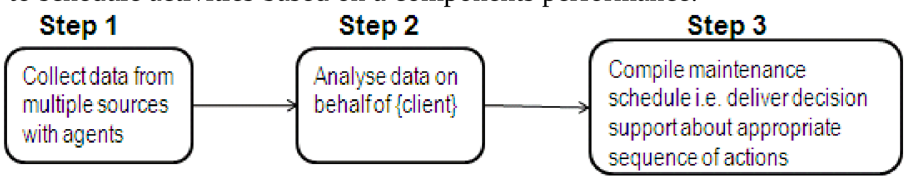

Figure 2 Steps for analysis of performance data

Step 1 includes: Identification of influencing factors; identification of relevant formulas; measuring appropriate variables; extraction of data from Data Warehouse and formatting of this data; and interpolate for missing values in MatLab. Step 2 includes: perform data mining procedures in Oracle data miner; analyse results to choose best data mining technique. Step 3: Utilise predicted values to schedule maintenance.

The experimental procedure used is as follows:
Formulation of Problem: Here the appropriate performance level at which maintenance should be performed, so as to prevent failure, is found.
Sampling Procedure: Take a component which is being used regularly, as per its design specification, for which maintenance is performed. Identify performance metrics for component; identify all possible failure types.
Collection of Data: Data is collected in a Data Warehouse, every 15 minutes. Data is validated by a simple check to ensure it is within expected ranges. Interpolation is carried out for gaps in data.
Analysis of Data: Using multiple analysis techniques, performance is predicted.

5 IMPLEMENTATION – SOLAR CIRCUIT PERFORMANCE PREDICTION

This methodology for performance based maintenance scheduling is implemented for a solar circuit which is present in the ERI. For this implementation a number of assumptions are made:

- Performance equation of one solar panel can be applied to circuit of solar panels
- A performance of <60% is regarded as a failure with respect to user comfort and energy efficiency

The method for evaluating the performance of these solar panels is to use the following equation which is provided by Vitosol in the user manual for a flat plate solar panel:

$$\pi = \pi_{\circ} - \frac{k_1\,\Delta T}{E_g} - \frac{k_2\,\Delta T^2}{E_g} \qquad (1)$$

Here π = efficiency, $\pi_{\circ}$ = optical efficiency, k1 = heat loss coefficient 1, k_2 = heat loss coefficient 2, ΔT = difference in temperature between fluid inside and ambient temperature outside, E_g = specific thermal heat capacity. A number of the values are specified by the manufacturer, as can be seen in Table 2.

Table 2: Flat plate collector values

Variable	Value	Units
$\pi_{\circ}$	84	%
k_1	3.36	$W/(m^2K)$
k_2	0.013	$W/(m^2K^2)$

The manufacturers specify the efficiency as between 0.5 – 0.68. This provides a range for determining if the collectors are operating at the expected range.

ΔT refers to the difference between the absorber surface and the ambient outside temperature. The varying temperature of the absorber surface is difficult to record, as 'an average value can only be determined from a measured temperature distribution', so where ΔT is specified, the following formula is used (Eicker 2003):

$$\frac{T_{Fin} + T_{Fout}}{2} - T_{outside} = \Delta T \qquad (2)$$

This can be used as long as the flow rate is not too low. The values for T_{fin}, T_{fout} and $T_{outside}$ are retrieved for the Data Warehouse; they are collected using wireless sensors and integrated with the building's BMS data.

Equation 1 is applied to historical and real time data which is stored in the data warehouse. By applying the equation, the performance of the solar panels is determined. This value is then checked against the expected values provided by the manufacturer as stated above and 1 or 0 is assigned to each row of data which will be analysed. With 1 indicating that maintenance is required, as the performance level is less than 60%, and 0 indicating that maintenance is not required, i.e. the performance level is greater than 60%.

Using many variables, which physically affect the performance of the solar circuit, data analysis is performed for the four techniques described above. This predicts the performance of a component, given the effect of the influencing variables.

6 DATA MINING RESULTS

6.1 Introduction

To understand the results appropriately, it is necessary first to introduce a number of terms. Firstly attribute importance refers to the variables affect on the predictive power of the data mining technique. It allows for the reduction in the number of variables used, which is helpful as too much information can reduce the effectiveness of the data mining techniques.

Secondly we introduce Receiver Operating Characteristic (ROC) curves. The horizontal axis of an ROC graph measures the false positive rate as a percentage. The vertical axis shows the true positive rate. The top left hand corner is the optimal location in an ROC curve, indicating high true-positive rate versus low false-positive rate. The area under the ROC curve measures the discriminating ability of a binary classification model. The area under the curve measure is especially useful for data sets with unbalanced target distribution (Oracle 2005). The threshold is a default value assigned by the Data mining program. It is used to manage the trade-off between true positive rate and false positive rate.

Next we introduce lift. It measures how well the model improves predictions over using a random value. You can graph the lift as either Cumulative Lift (the default) or as Cumulative Positive Cases. The x-axis of the graph is divided into quantiles (Oracle 2005).

6.2 Discussion of Results

The first analysis that was performed was to find out which attributes had the most affect of the predictive power of the technique. In figure x, the attribute importance graph, we can see that the total radiation has a large influence on the amount of heat generated as we would expect. We also see that the temperature of the water flowing from the calorifier through to the solar panel is the second biggest influence on the heat generated. This is somewhat surprising, but can be attributed to the occasions when valve VS2 is open and water is flowing from the Calorifier to the solar panel circuit.

Contradictory to this is that the temperature measured closer to the inlet of water for the solar panels is ranked only 6th in importance to the rate of heat generated. Also, it is surprising that the temperature measured closer to the output of the solar panel has a higher importance compared to the temperature measured by the heat meter. A reason for this may be that the specified positions of the temperature sensors are not correct.

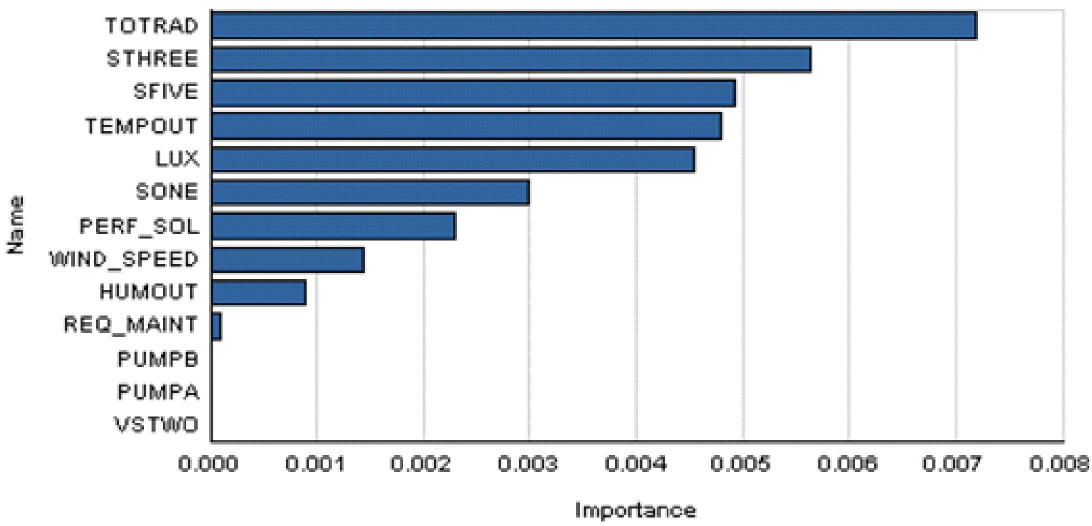

Figure 3 Attribute Importance

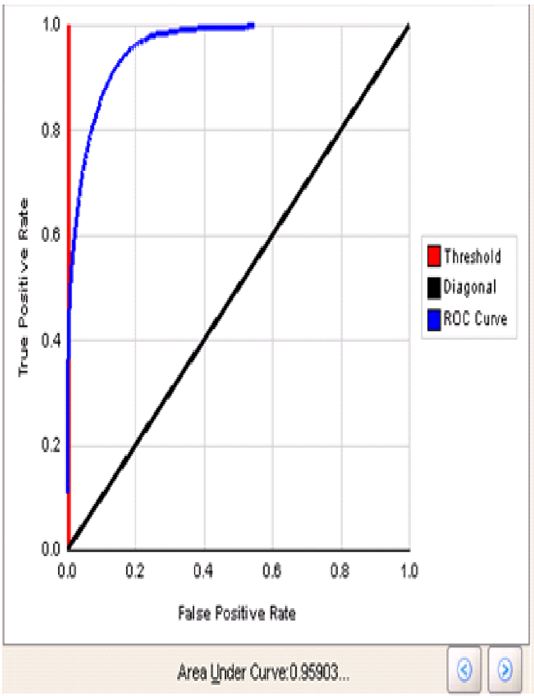

Figure 4 ROC curve for GLM

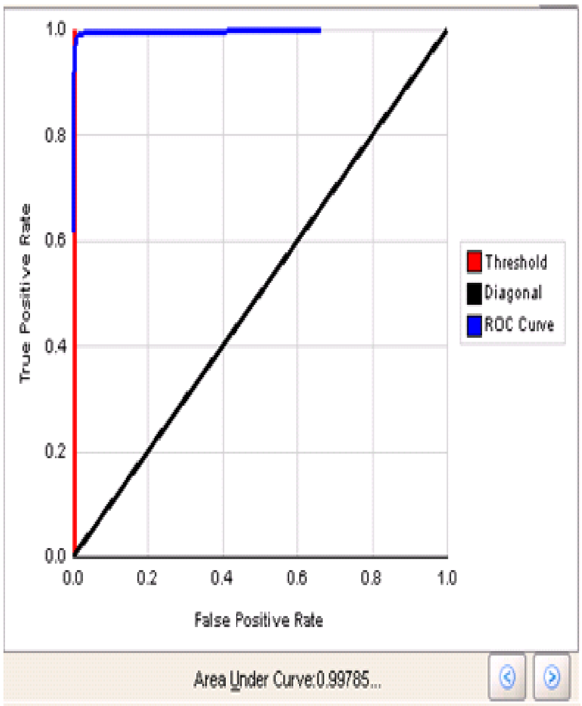

Figure 5 ROC curve for SVM

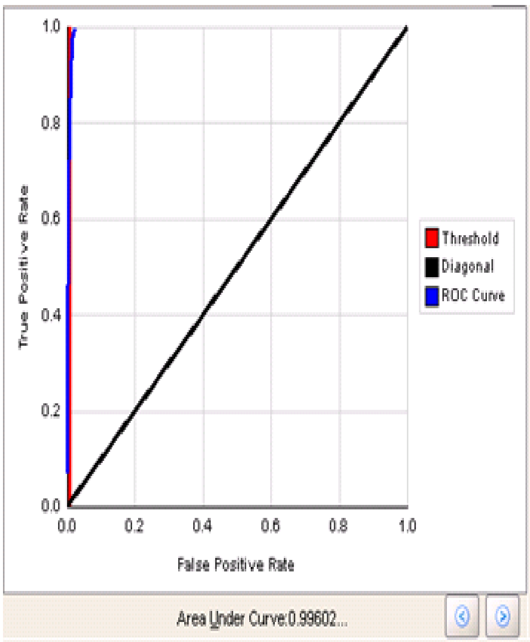

Figure 6 ROC curve for Naïve Bayes

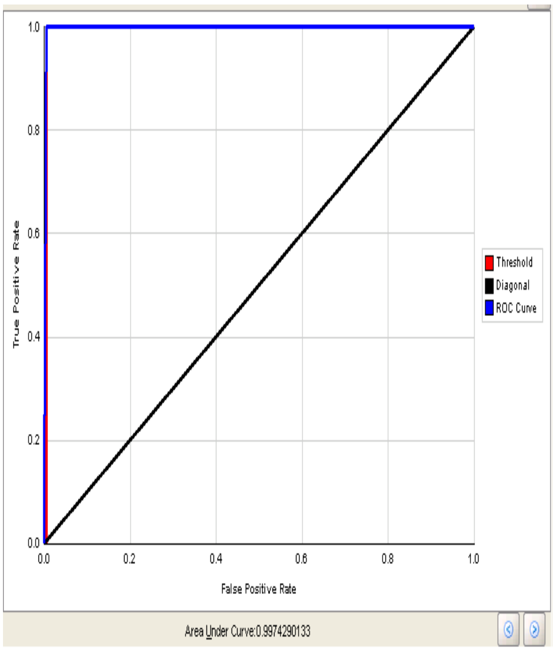

Figure 7 ROC curve for Decision Trees

For the GLM model, a predictive confidence of 4.51% is achieved. This result is very poor and indicates that the model is not much better than a naïve model. This can be attributed to the fact that the data being analysed are not linear. When the confusion matrix is viewed, Table 1, it is seen that this model is not usable for predicting performance.

Table 1: Confusion Matrix for GLM

Confusion Matrix	*0*	*1*
0	595,032	0
1	17,263	815

For the Support Vector Machines (SVM) model, a predictive confidence of 97.55% is achieved. This is an improvement on the GLM model, but when the confusion matrix is looked at in greater detail, it can be seen that there is still a substantial amount of times when it predicts that maintenance is required when it is not.

Table 2: Confusion Matrix for SVM

Confusion Matrix	*0*	*1*
0	588,995	6,037
1	260	17,818

For the Naïve Bayes model, a similar predictive confidence is calculated, 97.44%, and the confusion matrix shows a larger number of false positives, i.e. it predicts 13,295 times that maintenance is required when, in fact, it is not.

Table 3: Confusion Matrix for Naïve Bayes Model

Confusion Matrix	*0*	*1*
0	581,737	13,295
1	50	18,028

For the decision trees model, a predictive confidence of 99.04% was calculated. This model proved to be the most successful of the models, in that its confusion matrix showed the least amount of false positives and negatives combined.

Table 4: Confusion Matrix for Decision Trees Model

Confusion Matrix	*0*	*1*
0	591,986	3,046
1	81	17,997

7 IMPLEMENTATION OF DECISION SUPPORT MODEL

The results from data mining techniques can then be used to inform during a decision making process with regards to a choice between maintenance and renovation. The following diagram is an example from a Decision Model Framework (DMF), (Yin et al. 2011).

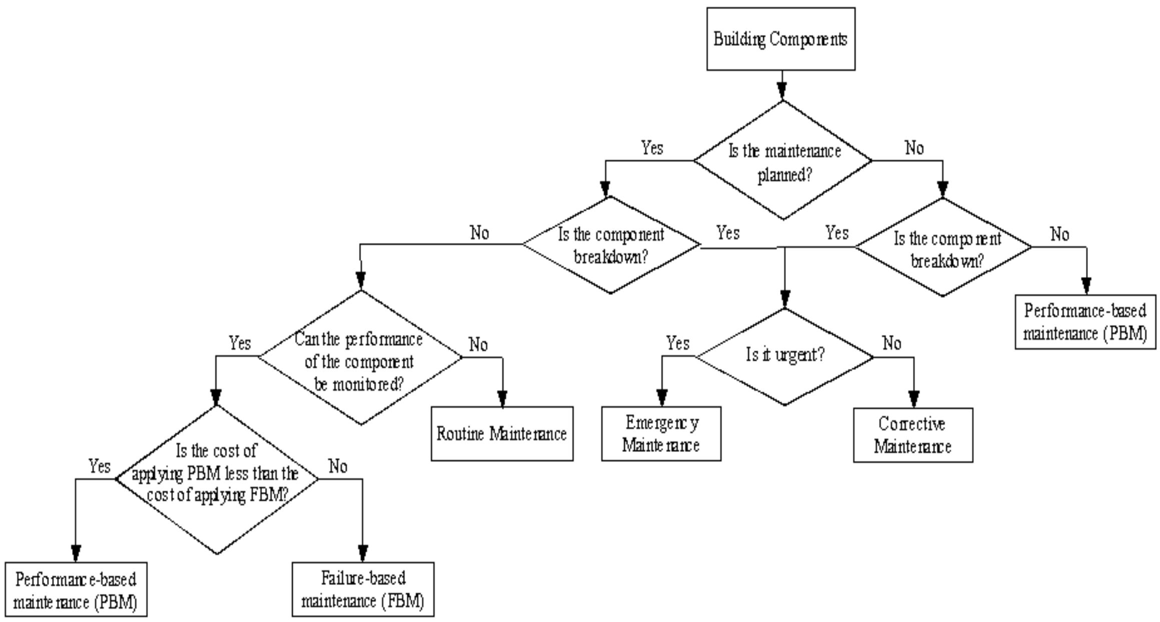

Figure 8: Decision Model Framework for Building Components

For example, in Figure 8, one would utilise the predicted value for the solar panel performance to decide if maintenance should be performed.

8 CONCLUSIONS

In conclusion, from the results presented in this paper, it is evident that a Decision Trees model is the most appropriate technique for predicting the performance of a solar panel circuit. The decision trees, Naïve Bayes and Support Vector Machines techniques proved to have a high predictive confidence but when the ROC curves are viewed, it is clear that the data is too conforming, as the ROC curves disappear mostly into the left hand corner of the graph. This is a point of note from these research activities and indicates that a higher percentage of the data should be kept for testing the model (at present it is 60% build, 40% testing).

There are a number of tasks which lead on from this research paper, firstly a second method of predicting the performance will be utilised. This will involve extracting curves which represent the performance of the solar circuit from the data and utilising these curves to schedule maintenance.

Also, it is necessary to integrate these results into the IT tool created to support the Decision Making Framework as per the work (Yin et al. 2011).

REFERENCES

E. Tobin, H. Yin, and K. Menzel (2010a), "Methodology for Maintenance Management Utilising Performance Data," in *eWork and eBusiness in Architecture, Engineering and Construction*, pp. 331-338.

H. Yin**,** K. Menzel (2011), "Decision Support Model for Building Renovation Strategies," *International Conference on Building Science and Engineering (ICBSE)* 2011, Venice, Italy, April. *(accepted)*

EC (European Commission) (2010), "Energy: Energy Efficiency in Buildings", http://ec.europa.eu/energy/efficiency/buildings/buildings_en.htm.

EN 13306: 2001 (2001), Maintenance terminology, *CEN (European Committee for Standardization),* Brussels.

M. Schwabacher and K. Goebel (2007), "A Survey of Artificial Intelligence for Prognostics," *in AAAI Fall Symposium,* Arlington VA.

J. Gu, N. Vichare, T. Tracy, and M. Pecht (2007) "Prognostics Implementation Methods for Electronics," *Reliability and Maintainability Symposium*, pp. 101-106. Orlando, FL

N. Vichare and Pecht (2006), M.; "Prognostics and Health Management of Electronics," *IEEE Transactions on Components and Packaging Technologies*, Vol. 29, No. 1, March 2006. pp. 222–229

Oracle (2005); "Oracle Data Mining Concepts, 11g Release1 (11.1)", *Part number B28129-04*

E. Tobin & K. Menzel (2010b), Optimal maintenance activities through collaborative work environment, 2010, *ForumBauInformatik*, No. 23, September 2010. pp. x-x

U. Eicker (2003); *Solar Technologies for Buildings*, ISBN 0-471-48637-X