Energy Consumption Prediction by Using Machine Learning

Nitesh Kushwaha[1], Akhilesh A. Waoo[2]

[1,2]Department of Computer Science and Engineering. AKS University, Satna, Madhya Pradesh, India

Abstract:
Energy consumption prediction is a critical task in today's world, where sustainable energy management and resource optimization are of paramount importance. This abstract presents a machine learning-based approach for accurately predicting energy consumption. By leveraging historical data and various predictive features, our model aims to provide accurate forecasts, enabling better energy resource allocation and efficient energy management.
In this study, we employ a diverse dataset comprising information such as weather conditions, time of day, building occupancy, and energy consumption records. We explore the use of several machine learning algorithms, including linear regression, decision trees, random forests, and neural networks, to find the most suitable model for energy consumption prediction.
The evaluation of our models is carried out through cross-validation and performance metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). We also investigate the feature importance to gain insights into the factors influencing energy consumption.
Our findings demonstrate the effectiveness of machine learning in accurately predicting energy consumption, with the potential to improve energy efficiency, reduce costs, and minimize environmental impacts. This research contributes to the broader efforts of smart energy management and paves the way for a more sustainable and responsible use of energy resources.

Keywords: Building energy management system Machine learning Microsoft Azure Energy consumption Prediction

Introduction:
Energy consumption prediction is a critical and challenging task in various sectors, including utilities, manufacturing, transportation, and even residential households. Accurate predictions of energy consumption play a crucial role in optimizing resource allocation, reducing costs, and promoting sustainability. Machine learning, a subset of artificial intelligence, has emerged as a powerful tool for tackling this issue.
Machine learning techniques leverage historical data, various features, and algorithms to model and predict future energy consumption patterns. These predictions can assist organizations and individuals in making informed decisions, such as optimizing energy usage, implementing energy-efficient technologies, and managing energy resources more effectively.
This paper aims to explore the application of machine learning in predicting energy consumption, highlighting its importance and the methodologies employed. We will discuss the following key aspects:

1. **Problem Statement:** Defining the problem of energy consumption prediction and its significance in different domains. This section will outline the challenges associated with forecasting energy consumption accurately.
2. **Data Collection:** The foundation of any machine learning model is the data it uses. We will delve into the various data sources, such as smart meters, weather data, and occupancy patterns, that are crucial for energy consumption prediction.
3. **Feature Engineering:** Extracting meaningful features from the collected data is essential for creating predictive models. We will discuss the process of feature engineering and how it impacts the model's performance.
4. **Model Selection:** Evaluating different machine learning algorithms and selecting the most appropriate one for energy consumption prediction. Commonly used models include linear regression, decision trees, neural networks, and time series forecasting methods.
5. **Training and Validation:** The process of training the machine learning model on historical data and validating its performance using different evaluation metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE).
6. **Challenges:** Discussing the challenges and limitations associated with energy consumption prediction using machine learning, such as data quality, seasonality, and external factors like policy changes and economic conditions.
7. **Applications:** Highlighting the real-world applications of energy consumption prediction, including load forecasting for utility companies, demand-side management in smart grids, and energy-efficient building management systems.
8. **Future Directions:** Exploring emerging trends and future research directions in the field of energy consumption prediction. This section will touch upon the integration of IoT devices, advancements in deep learning, and the potential for enhancing predictive accuracy.

By providing a comprehensive overview of energy consumption prediction through machine learning, this paper aims to promote a better understanding of the methodologies, challenges, and potential benefits of this innovative approach. As energy conservation and sustainability continue to be at the forefront of global concerns, accurate energy consumption prediction can be a driving force for positive change in the way we manage and utilize energy resources.

Machine learning prediction methodology

Energy consumption prediction using machine learning typically follows a well-defined methodology to develop accurate models. Here's a general outline of the steps involved in this process:

1. **Data Collection**:
- Gather historical energy consumption data. This data may include information about time, location, weather conditions, building characteristics, and energy usage.

2. **Data Preprocessing**:
- Clean the data by handling missing values, outliers, and errors.
- Transform and normalize the data if necessary. For instance, you might convert timestamps into numerical features, or scale the data to have zero mean and unit variance.

3. **Feature Engineering**:
- Extract or create relevant features that can help the model make accurate predictions. This may involve adding weather data, day of the week, or other contextual information to the dataset.

4. **Data Splitting**:
- Split the dataset into training, validation, and test sets. Typically, 70-80% of the data is used for training, 10-15% for validation, and 10-15% for testing.

5. **Model Selection**:
- Choose the appropriate machine learning algorithm for your problem. Common choices include linear regression, decision trees, random forests, support vector machines, and neural networks. The choice may depend on the complexity of the problem and the size of the dataset.

6. **Model Training**:
- Train the selected model using the training dataset. The model will learn the relationships between the features and energy consumption patterns.

7. **Hyperparameter Tuning**:
- Optimize the hyperparameters of the model using the validation dataset. Techniques like grid search or random search can be used to find the best hyperparameters.

8. **Model Evaluation**:
- Evaluate the model's performance using the test dataset. Common evaluation metrics for regression tasks in energy consumption prediction include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

9. **Model Interpretation (Optional)**:
- Depending on the model used, you may want to interpret its results. Some models, like decision trees or linear regression, provide feature importance or coefficients that can help understand which features contribute the most to energy consumption.

10. **Deployment**:
- Once you are satisfied with the model's performance, you can deploy it to make real-time predictions or automate energy management decisions.

11. **Monitoring and Maintenance**:
- Continuously monitor the model's performance and retrain it if necessary. Energy consumption patterns can change over time due to various factors, so it's essential to keep the model up to date.

12. **Feedback Loop (Optional)**:
- Collect feedback and additional data to improve the model over time. User feedback and new data can help refine the predictions and adapt to changing conditions.

13. **Documentation and Reporting**:
- Document the entire process, including data sources, preprocessing steps, model architecture, and results. This documentation is crucial for knowledge sharing and compliance.

Energy consumption prediction is an important application of machine learning in the context of energy management, demand forecasting, and resource optimization. The choice of algorithms and the complexity of the model will depend on the specific requirements and constraints of the problem. Additionally, domain expertise is valuable for feature selection and understanding the significance of the results.

Normality testing of dataset

Normality testing is a statistical procedure used to determine whether a dataset follows a normal distribution, which is a key assumption in many statistical methods, including some machine learning algorithms. To perform normality testing on your dataset for energy consumption prediction, you can follow these steps:

1. **Data Collection**: Collect your dataset, which should include historical energy consumption data and any other relevant features that you plan to use for prediction.
2. **Data Preprocessing**: Clean and preprocess your data by handling missing values, outliers, and any other data quality issues.
3. **Select Your Variable of Interest**: Identify the variable that represents energy consumption, as this will be the one you want to test for normality.
4. **Visual Inspection**: Start by creating visualizations such as histograms and Q-Q plots to get an initial sense of the data's distribution. This can help you detect major departures from normality.

Statistical Normality Tests:

a. **Shapiro-Wilk Test**: The Shapiro-Wilk test is a common statistical test for normality. It assesses whether the data follows a normal distribution. In Python, you can use the **scipy.stats.shapiro()** function to perform this test.
b. **Anderson-Darling Test**: The Anderson-Darling test is another test for normality that takes into account the shape of the distribution. You can use the **scipy.stats.anderson()** function for this test.
c. **Kolmogorov-Smirnov Test**: The Kolmogorov-Smirnov test can be used to compare your data's distribution to a normal distribution. You can use the **scipy.stats.kstest()** function for this test.

Interpret the Results:

a. If the p-value from the normality test is less than your chosen significance level (e.g., 0.05), you would reject the null hypothesis, indicating that your data is not normally distributed.
b. If the p-value is greater than your chosen significance level, you would fail to reject the null hypothesis, suggesting that your data may be normally distributed.

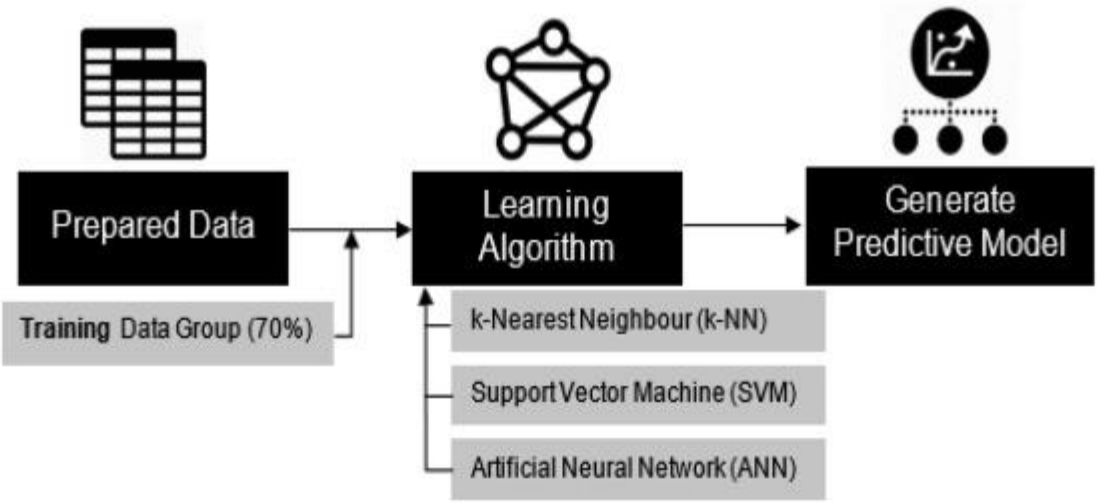
5. **Consider Data Transformations**: If your data is found to be significantly non-normal, you may consider applying data transformations (e.g., logarithmic or Box-Cox) to make the data more normally distributed.
6. **Machine Learning Model Selection**: Depending on the results of the normality test and the nature of your data, you can choose a machine learning algorithm that is less sensitive to the normality assumption or apply appropriate data transformations to mitigate non-normality.

Remember that normality is just one assumption among many in the context of machine learning, and the impact of deviations from normality may vary depending on the specific algorithm you are using. In some cases, machine learning algorithms are robust to departures from normality, while in others, data transformations or different modeling approaches may be necessary.

Model Development (Training)

This research used a supervised machine learning methodology to predict energy consumption. After data was prepared, it was then inputted into the learning algorithm. Different feature combinations were fed into the algorithm to generate a candidate for the predictive model. Before using the data to create and train the model, data partitioning was done to separate the data into two groups – a training group and a testing group.

The predictive modeling for this research used a classification method to predict discrete variables instead of regressive prediction. As Azure ML does not have k-Nearest Neighbour and Artificial Neural Network for classification, the modeling function in Caret R package was utilised for all prediction to ensure uniform execution. Three types of machine learning algorithm were used for this research which were Artificial Neural Network (ANN-MLP), k-Nearest Neighbour (k-NN), and Support Vector Machine (SVM-RBF).



Normality testing of dataset

Table 1. Skewness for each tenant using aggregate data.

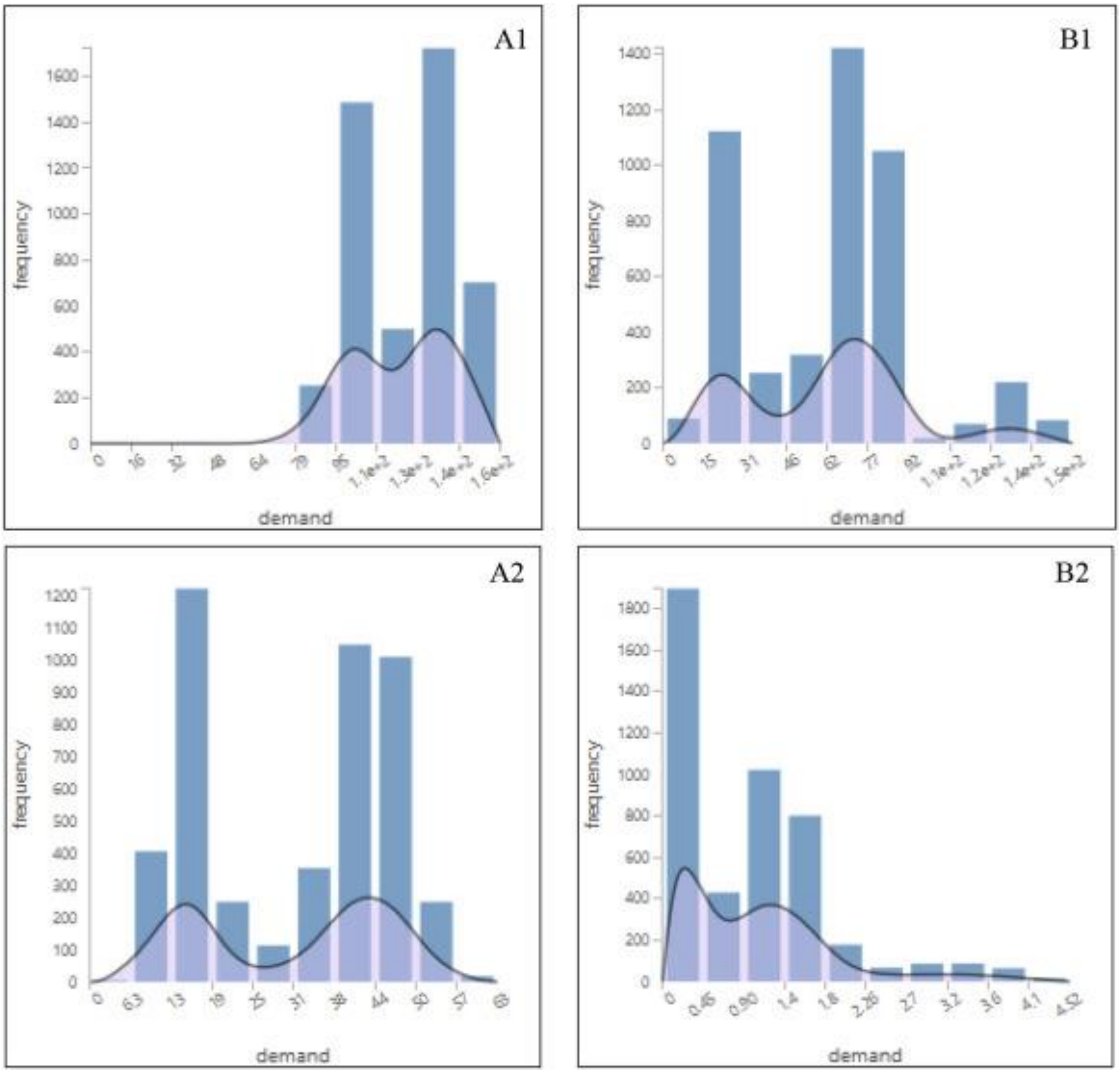
Tenants	**Skewness**			
	Power Factor	**Current**	**Voltage**	**Demand**
A1	−0.457296	−0.17541	−0.116392	−0.279034
A2	−1.564851	−0.105746	−0.054989	−0.182159
B1	−1.877798	−0.666165	−0.310973	0.282481
B2	1.735751	0.202681	−0.714503	1.267578

Table 2. Kurtosis level for each tenant using aggregate data.

Tenants	**Kurtosis**			
	Power Factor	**Current**	**Voltage**	**Demand**
A1	6.026131	−1.648018	−0.45877	−1.053977
A2	2.839891	−1.763198	−0.768862	−1.584641
B1	2.062267	−1.011822	0.469322	−0.126043
B2	3.472321	−1.403968	4.623413	1.824909

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com



Imputation of missing data

The study of missing data was conducted inside Azure ML studio whereby Summarize Data module was utilised to determine the amount of missing data and to reveal the rows which have missing data. The diagnosis of missing data was also conducted via observation on the value of other attributes in the same row. shows a summary of the analysis.

Performance evaluation and comparison

Subsequently, after model training and testing, the predictive model generated was compared in terms of performance between algorithms for each tenant. Initially, the result of the testing was observed by comparing the performance of the methods for individual tenants. The comparison table is as shown in

Table 5. Performance evaluation of test prediction for all tenants using trained SVM, k-NN and ANN model.

Tenant	Method	RMSE	NRMSE (%)	MAPE (%)
A1	k-NN	5.0025748	4.06	3.02

Tenant	Method	RMSE	NRMSE (%)	MAPE (%)
	SVM	4.7506789	3.85	2.76
	ANN	8.874015	7.19	5.02
A2	k-NN	3.6548885	11.46	9.98
	SVM	3.5898263	11.25	9.38
	ANN	4.540988	14.23	14.16
B1	k-NN	14.934312	23.87	15.43
	SVM	16.0690844	25.69	12.09
	ANN	20.63566	32.99	28.00
B2	k-NN	0.5439403	55.87	48.75
	SVM	0.5558279	57.09	43.97
	ANN	0.547152	56.20	60.62

Tenants	Number of missing data				Total number of missing data
	P. Factor	Current	Voltage	Demand	
A1	0	0	0	1	1
A2	0	0	0	3	3
B1	0	0	0	1	1
B2	0	0	0	171	171

Result Generation

Predicting energy consumption using machine learning can be a valuable application in various domains, such as smart grid management, energy efficiency, and cost optimization. To generate predictions for energy consumption, within follow these steps:

1. **Data Collection**:

Gather historical energy consumption data. This data may include information such as date and time, weather conditions, building characteristics, and other relevant variables. You may obtain this data from utilities, sensors, or publicly available datasets.

2. **Data Preprocessing**:

Clean and preprocess the data. Handle missing values, outliers, and data quality issues.
Convert categorical variables into numerical format using techniques like one-hot encoding or label encoding.

3. **Feature Selection/Engineering**:

Identify relevant features that may impact energy consumption. This could include factors like

temperature, humidity, time of day, and seasonal patterns.
Create new features that might capture hidden patterns or correlations.

4. **Data Splitting**:
Split the data into training and testing datasets. The training data will be used to train your machine learning model, and the testing data will be used to evaluate its performance.

5. **Model Selection**:
Choose an appropriate machine learning model for your energy consumption prediction task. Common models include linear regression, decision trees, random forests, support vector machines, and neural networks.

6. **Model Training**:
Train the selected model using the training dataset. The model will learn the patterns and relationships within the data.

7. **Model Evaluation**:
Evaluate the model's performance using the testing dataset. Common evaluation metrics for regression tasks include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

8. **Hyperparameter Tuning**:
Fine-tune your model by adjusting hyperparameters. This can be done using techniques like grid search or random search.

9. **Model Deployment**:
Once you are satisfied with your model's performance, deploy it for making real-time predictions. This may involve integrating it into an application, a web service, or a dashboard.

10. **Monitoring and Maintenance**:
Continuously monitor the model's performance in a production environment and retrain it periodically with new data to ensure it remains accurate.

11. **Interpretation**:
Understand how the model is making predictions. Interpretability is crucial in energy consumption prediction, especially when it's used for decision-making.

12. **Forecasting**:
Use the trained model to make energy consumption predictions for future time periods. Incorporate weather forecasts or other relevant data for accurate future predictions.

Visualization and Reporting:
Present the results of your energy consumption predictions in a user-friendly manner, which could include charts, graphs, or reports for stakeholders.
Keep in mind that the choice of the machine learning algorithm and the success of your model will depend on the specific characteristics of your energy consumption dataset. Additionally, domain knowledge and understanding of the factors affecting energy consumption are crucial for feature selection and model interpretation.

Future Scope

Predicting the future scope of a specific application or field within Machine Learning is a challenging task, as it depends on various factors such as technological advancements, market trends, and societal needs. However, I can provide some general insights on the future scope of Machine Learning result generation:

1. **Increased Automation**: Machine Learning will continue to play a pivotal role in automating tasks that were traditionally done by humans. This includes generating reports, insights, and recommendations from large datasets. As ML algorithms become more sophisticated, they will be capable of producing more accurate and valuable results.
2. **Personalization**: The demand for personalized content and recommendations will drive the future scope of ML result generation. Companies and service providers will increasingly use ML to tailor content and services to individual preferences, creating a better user experience.
3. **Natural Language Generation (NLG)**: NLG is an application of ML that converts structured data into human-readable text. This has applications in generating reports, news articles, and more. As NLG algorithms improve, we can expect to see wider adoption in various industries.
4. **Healthcare**: Machine Learning will continue to impact healthcare by assisting in the generation of medical reports, diagnosis, and treatment recommendations. ML models can analyze patient data to provide more accurate and timely results.
5. **Financial Services**: In the financial sector, ML can help in risk assessment, fraud detection, and investment recommendations. The future will likely see ML models playing a more significant role in generating financial reports and forecasts.
6. **Content Generation**: ML will continue to evolve in the realm of content generation, including the creation of art, music, and even computer code. Generative models like GPT-3 and successors will become more sophisticated and capable.
7. **Natural Language Understanding**: As ML models become better at understanding and interpreting human language, they will play a key role in applications like chatbots, virtual assistants, and customer service. This will enhance the user experience and automate routine tasks.
8. **Ethical Considerations**: The ethical use of ML in result generation will be a significant concern in the future. The development of guidelines and regulations for responsible AI and ML use will shape the scope of ML result generation.
9. **Interdisciplinary Applications**: Machine Learning will increasingly intersect with other fields like robotics, healthcare, and autonomous vehicles. This will create opportunities for innovative applications of ML in generating results that solve complex problems.
10. **Continuous Innovation**: Machine Learning is a rapidly evolving field. The future scope will be shaped by continuous innovation, breakthroughs in AI research, and the development of more powerful and efficient algorithms.

It's important to note that the scope of Machine Learning result generation is vast and continually evolving. To stay relevant in this field, professionals need to adapt to emerging technologies, acquire new skills, and stay updated with the latest developments in the ML and AI communities.

Conclusion

Predicting conclusions from machine learning results can be a valuable aspect of data analysis. Here's how you can generate conclusions based on machine learning results:

1. **Understand the Problem**: Before generating conclusions, ensure that you have a clear understanding of the problem you are trying to solve and the dataset you are working with. Understand the context and goals of your analysis.

2. **Data Preprocessing**: Ensure that your data is clean and preprocessed. This includes handling missing values, encoding categorical variables, and scaling/normalizing data as necessary.
3. **Feature Importance**: Use feature selection or extraction techniques to identify the most important variables that contributed to the machine learning model's predictions. This can help you understand the key factors driving the results.
4. **Model Evaluation**: Assess the performance of your machine learning model using appropriate evaluation metrics. This could include metrics like accuracy, precision, recall, F1-score, or area under the ROC curve, depending on the type of problem (classification or regression).
5. **Interpretability**: Depending on the complexity of your model, use techniques like feature importance plots, SHAP (SHapley Additive exPlanations) values, or partial dependence plots to interpret how individual features affect the model's predictions.
6. **Generalization**: Evaluate whether your machine learning model can generalize well to unseen data. Cross-validation and validation curves can help in assessing this aspect.
7. **Domain Knowledge**: Combine your machine learning results with domain expertise. It's essential to validate the conclusions from the model with what makes sense in the real-world context.
8. **Uncertainty and Confidence Intervals**: Consider estimating uncertainty in your predictions. Confidence intervals or uncertainty estimates, such as prediction intervals in regression, can help you understand the range of possible outcomes.
9. **Bias and Fairness**: Assess and address any biases in your data or model. Ensure fairness and ethical considerations are taken into account in the conclusions you draw.
10. **Communication**: Clearly communicate your conclusions in a way that is understandable to both technical and non-technical stakeholders. Visualization tools and storytelling techniques can be helpful.
11. **Next Steps**: Suggest actionable insights or recommendations based on the conclusions. What should be done in light of the machine learning results? What decisions can be made, or what additional data should be collected for further analysis?
12. **Validation**: Validate your conclusions by applying them to new, unseen data if possible. This helps ensure that your conclusions are not overfit to the training data.
13. **Documentation**: Document your methodology, findings, and any limitations of the analysis. This is important for transparency and reproducibility.

In summary, generating conclusions from machine learning results involves a combination of statistical analysis, data interpretation, domain expertise, and effective communication. It's important to be thorough and rigorous in your approach to draw meaningful and actionable insights from your machine learning models.

References:

1. Jones, Andy (2009, 20 June). Lessons not learned on data disposal, digital investigation 6(2009) - Elsevier, p.3–7
2. Jones, Andy (2009, November). “Recycling more than your IT equipment” - Network Security, Volume 2009, Issue 11, Elsevier, p. 8-9.
3. Joumaa Chibli, Kadry Seifedine, (2012).Green IT: Case Studies. 2012 International Conference on Future Energy, Environment and Materials. Energy Procedia 16, p. 1052 –1058. Elsevier
4. Karger, P.A, (2008, September-October). I/O for virtual machine monitors Volume: 6 , Issue: 5.

pp, 16 – 23. Security & Privacy, IEEE.
5. Kaufman, Lori M. (2009). Data Security in the World of Cloud Computing. Co published by the IEEE computer and reliability societies. p, 61-64. IEEE security and privacy.,
6. Kavanagh, Ronan (2009, March). IT Virtualization Helps to Go Green. Information Management volume 19. Issue 2. ProQuest Central
7. Kohli, R., and Grover,V. (2008). "Business Value of IT: An Essay on Expanding Research Directions to keep up with the Times," Journal of the Association for Information Systems, Volume 13, Number 1, p. 23-38.
8. Krikke Jan.(2008). Recycling e-Waste: The Sky Is the Limit. IT Pro, Published by IEEEComputer Society.
9. Kwon Young Chul, Lee Sang Won, and Moon Songchun (2006). Personal Computer Privacy: Analysis for Korean PC Users. IWSEC 2006, LNCS 4266, p. 76–87- Springer
10. Lamb, John P (2009, April). The greening of IT: how companies can make a different for the environment. Upper Saddle River, NJ:IBM Press. Preface: page xxv.
11. Lamb, John P (2011). Green IT and use of Private Cloud Computing in South Africa. 8th International Conference Expo on Emerging Technologies for a Smarter World (CEWIT), 2011. IEEE.
12. Lechner Rich, (2007, 24 September). Using virtualization to boost efficiency. Tech Update, networkworld.com- Proquest Central. p-24
13. Lee. Young Choon and Zomaya Albert. Y, (2010, 19 March). Energy efficient utilization of resources in cloud computing systems. The journal of supercomputing. Springer
14. Leung Adrian, Shengb Yingli and Cruickshank Haitham (2007), The security challenges formobile ubiquitous services. Information security technical report 12 (2007) pp.162–171. Elsevier
15. Levy, Y., & Ellis, T. J. (2006). A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. Informing Science, Volume 9.
16. Li. Jianxin , Li. Bo, Wo. Tianyu, Hu. Chunming, Huai. Jinpeng, Liu. Lu and Lam. K.P. (2011, 11 May). Cyber Guarder: A virtualization security assurance architecture for green cloud computing . Future Generation Computer Systems 28(2012) p.379–390
17. Li, Qilin & Zhou Mingtian. (2011). The Survey and Future Evolution of Green Computing. ACM International Conference on Green Computing and Communications, IEEE 2011, pp p. 230-233.
18. Liang. Dung-Hai, Liang. Dong-Shong and Chang. Chun-Pin. (2012). Cloud Computing and Green Management. 2012 International Conference on Intelligent System Design and Engineering Application.
19. Liu Lu, Masfary Osama and Li Jianxin, (2011).Evaluation of Server Virtualization Technologies for Green IT. Proceedings of The 6th IEEE International Symposium on Service Oriented System Engineering (SOSE 2011), pp-79-84. IEEE Computer Society
20. Lombardi Flavio and Pietro Roberto Di, (2010). Secure virtualization for cloud computing.Journal of Network and Computer Applications 34 (2011) pp, 1113–1122.
21. Mahalingam. P, Jayaprakash. N and Karthikeyan. S, (2009). Enhanced Data Security Framework for Storage Area Networks. 2009 Second International Conference on Environmental and Computer Science - IEEE Computer Society, p.105-110.
22. Mathieson, (2007). Don't be green. Infosecurity Volume 4, Issue 6, September 2007-Elsevier, p. 26–29.

23. McAlearney Shawna, (2007 , January 30). Dispose of IT Equipment Without Sharing Secrets. Retrieved from, http://www.cio.com/article/216257/Dispose_of_IT_Equipment_Without_Sharing_Secrets
24. McConnell. M, (2002). Information Assurance in the Twenty-First Century. Supplement to Computer. Security & Privacy 2002
25. McKinsey Inc. (2007). Reducing U.S. Greenhouse Gas Emissions: How Much at What Cost? Green IT: Corporate Strategies.
26. McKnight. Walter L. (2002, July). "What is Information Assurance?" The Journal of Defense Software Engineering. p. 4-6.
27. Mell.P and Grance. T, NIST. (2009, June). Accessed from Security Guidance for Critical Areas of Focus in Cloud Computing V2.1.