National College of Ireland

Water, Gas & Electricity Consumption Behaviour Forecasting

MSc Research Project
MSc Data Analytics

Abhilash Anil Chavan
Student ID: X17170907

School of Computing
National College of Ireland

Supervisor: Anu Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Abhilash Anil Chavan……

Student ID: X17170907……..……

Programme: MSc Data Analytics…… **Year:** 2018-19……..

Module: MSc Research Project…….……

Supervisor: Dr Anu Sahni…..……

Submission Due Date: 12th August 2019……..……

Project Title: Water, Gas & Electricity Consumption Behaviour Forecasting ……

Word Count: 5,300…… **Page Count** 15 ……..……..……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……

Date: 9th August 2019 ……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Water, Gas & Electricity Consumption Behaviour Forecasting

Abhilash Anil Chavan
X17170907

Abstract

Water, gas and electricity are precious resources available to us, and fresh drinking water has limited availability. Daily millions of households and industries are supplied with these resources and forecasting the consumption need would help the managing authorities regularize the working timetables for these supplies. This study is based on the hypothesis that there are peaks and drops in the consumers consumption behaviour and knowing those needs would help management in providing the consumers with a tailored plan for usage, reducing the cost of supply and labour. For accurately forecasting the consumption need LSTM is implemented as LSTM was designed for forecasting time series data, and for evaluation of the model RMSE values were checked rather than the accuracy as RMSE gives clear idea how far the predicted values are from actual values. The data used in this study is a time series data available from the smart meters at an 30min time interval. The implemented model achieved 0.21litres RMSE value and outperformed every other model. With this low error rate, the forecasted values would help authorities detect anomalies in the consumption and develop a plan of action.

1 Introduction

Many cities and countries all over the world are attempting for the sustainable development in order to effectively manage the available resources. This can be partly achieved by implementing better water, gas and electricity management regulations and optimum use of natural resources. One of the efficient ways of saving water and energy is to understand the consumption behaviour of all the consumers and predict the short and long-term water and energy needs. This not only helps in better management but makes authorities ready for draught like conditions and meet the needs of exponentially increasing population demands.

If the governing authorities of water, gas and electricity management can know the exact future demand of water and energy, they can take appropriate measures to understand the consumption behaviour, reduce loss through leakage pipes of water, and save enormous energy required to pump water when not required. Knowing the peak demand times and consumption requirements in advance would help reduce delays and save excessive electricity wasted on pumping. In (A, 2017) study found out that in North China Plain water consumption tripled in span of 24 years and is now suffering major water crises. (Parks and Mclaren, 2019) emphasized on how water shortages will be more common in coming years due to climate change, the study finds the various reasons leading to the day Zero in Cape Town, South Africa in 2017-2018 and gives a breakdown how the complete municipal water

network was shutdown due to lack of water management. To avoid these types of situations water, gas and electricity should be managed in a responsible manner, as there is very limited supply of fresh water and many water bodies are shared by different cities and countries, also production and transmission of electricity is very expensive. Misuse or unethical use of water and other energy resources will lead to severe water scarcity and can spark international wars. This benefits the authorities in following manner:

a) Improve working timetables for all sectors
b) Oversee the maintenance of pumps, pipes and transmission cables
c) Better understanding of water and energy demand time periods
d) Save unnecessary wastage of water and energy by implementing restrictions.

This study focuses primarily on forecasting the water, gas and electricity consumption demands to help authorities understand the consumer behaviour.

The objective of this study is to forecast the next months consumption demand based on the historical data available from the smart meters for water, gas and electricity.

This study is based on the hypothesis that there will be peaks and drops in the consumption behaviours instead of constant demand and thus requires continuous improvement in forecasting the consumers consumption behaviour.

This paper proposes the use LSTM for forecasting the consumers consumption behaviour of water, gas and electricity on a short-term basis with minimal error rate as compared to existing models and help the respective authorities to effectively manage resources.

The results in this study will be evaluated based upon the Root Mean Square Error values compared to the actual values rather than accuracy of predicted values, Also the recall of all metrics based on the obtained results. This paper is laid out in following manner: In section 2 related work carried out in the field of water consumption, forecasting and water consumption analysis and time series forecasting will be analysed and compared with this study. In section 3 covers research methodology of the study and section 4 focuses on design specification of this study. Section 5 presents the implementation carried out for the experiment and section 6 evaluates the results obtained from the experiment and also covers the detailed discussion of the results obtained and comparing it with related works results. Section 7 is dedicated for conclusion and limitations of the study carried out and future work is outlined.

2 Related Work

In order to attain sustainable development, many cities and countries conducted studies and implemented laws. Most studies were focussing on a small area and took only one aspect at a time, either water forecasting, or electricity consumption or anomaly detection. The prediction models built were based on probabilistic models such as hidden markov model as in (Liisberg *et al.*, 2016; and Leyli-Abadi *et al.*, 2018).

(Zucchini, Macdonald and Langrock, 2017) concentrated on analysis of discrete time series and different models that can be executed for the examination. The study concluded that there were very less discrete valued models were proposed. According to (Zucchini, Macdonald and Langrock, 2017) there were no notable group of models that could be fundamentally

minimal, simple, adequately flexible and provide a helpful variety of information types and also effectively available for the developers to utilize.

Melbourne Corporation did a similar forecasting back in 2000, but rather than just relying on the historical data available from smart meters the (Zhou *et al.*, 2000) included two more features for the prediction system. The (Zhou *et al.*, 2000) study implemented seven different models on a two decade long historical data along with temperature of the location and the rainfall of the location was also taken into the consideration. According to (Cutore *et al.*, 2008) all the forecasting models for water consumption were developed in a deterministic manner, and a realistic approach was needed to forecast more accurately. Inclusion of spatio-temporal variables and adding more features such as temperature and rainfall would paint a bigger picture. Although addition of these features would improve the performance of the model, but it adds unnecessary complexity as compared to using only historical data. Also fetching this new information would require more memory for the system and permissions from different departments.

SCEM-UA implemented by (Cutore *et al.*, 2008) improved the performance of the model and achieved 90% overall accuracy compared to traditional prediction models with the introduction of artificial neural networks, but as pointed by (Gers and Cummins, 1999) absence of forget gate in the architecture of the ANN prohibits it from resetting the learning mechanism, and thus does not perform well compared to the LSTM model. (Wei, Xingxing and Yanhong, 2008) suggested the use of SVM on top of the combination forecasting model to pick the best predicted values from different models. Although SVM performs well but SVM lacks the ability to predict the numerical values for time series data. SVM is best used for classification rather than forecasting time series data as mentioned by the (Kim, 2003) in their study how SVM is inferior to the ANN's accuracy.

A similar study performed by (Şen and Altunkaynak, 2009) built a fuzzy system model for forecasting the drinking water consumption, but took a novel approach to include the body temperature, weight of the people under study and their physical activity. These variables although enhanced the accuracy of prediction system but are not useful to predict other consumption areas such as electricity and gas consumption. (Aksela and Aksela, 2010) was one of the first studies to work on the smart meter readings for increasing the accuracy of the prediction models. Their approach was to first classify the consumption behaviours in different classes and then predict how the behaviour will change with time keeping temperature and rainfall as input variable along with the historical data.

But the data available for training their hidden markov model was an averaged monthly consumption values and had data coming from only households, this works in case of small areas but fails when meter readings from commercial as well as industrial places where peaks and drops vary significantly as compared to household usage. (Chamroukhi *et al.*, 2011) explains how a probabilistic model such as hidden markov model performs better than SVM and fuzzy systems but also highlights the drawbacks of HMMs, the assumptions made during forecasting and affect the performance of the model. (Adamowski *et al.*, 2012) on the other hand proposed combination of ANN's with the discrete wavelet transform for forecasting the consumers consumption behaviour for water, climatic variables were also included to find the correlation between consumption and surrounding environment. This study performed well in comparison with the multiple linear regression but lacks the accuracy what LSTM provides.

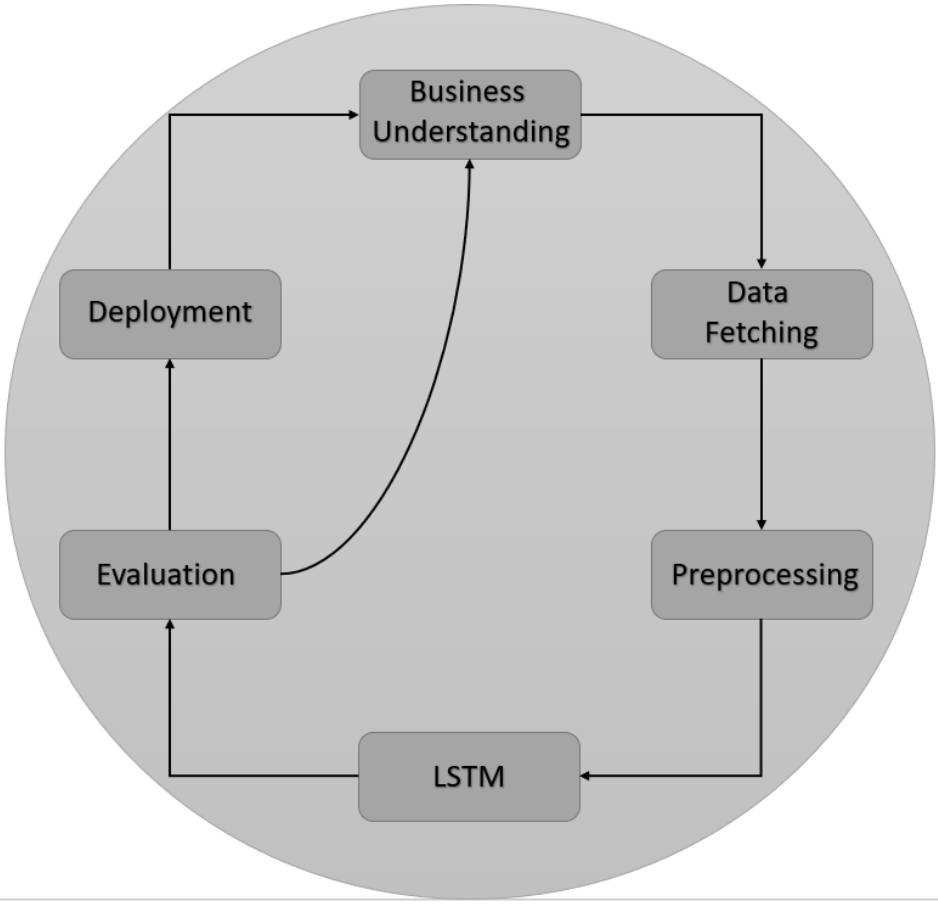
(Labeeuw and Deconinck, 2013) was smart grid project and showed promising results for forecasting short term needs for electricity. The study presented how markov chain can be used to prepare a probabilistic model, but the data used for training and testing was of only four months and that kind of data lacks seasonality of the year and might perform poorly when allowed to forecast for the year. The study involved very limited dataset and could have been tested for higher dataset for understanding how the model performs for the long term prediction. (Kwac, Flora and Rajagopal, 2014) proposed adaptive K-means algorithm for the forecasting of electricity consumption. The readings were fetched from smart meters on an hourly basis, algorithm was efficient enough to cluster the various consumption behaviours, but prediction model based on those clusters were not taking the increasing demand into consideration and thus did not seem viable for forecasting model.
(Hokimoto and Shimizu, 2014) was amongst the first studies to propose the hidden markov model due to their reliability and accuracy, the hidden markov model was facing the non-stationary covariates present in the dataset. (Dias and Ramos, 2014) proposed dynamic grouping of the electricity consumption, this study helped in synchronization of the various markets and bring the market on a standard level for business, but did not focus on improving the accuracy of the prediction model. (Gagliardi *et al.*, 2017) compared both the hidden Markov model and Non-Homogeneous hidden Markov model, the study concluded how NHMM is better than HMM's for forecasting the time series data, but still doesn't surpass the performance and accuracy of the LSTM models for forecasting the time series data, other models drastically fail with the presence of missing values and outliers, but in same situations LSTM outperformed every other model giving consistent results.

3 Research Methodology

To carry out the study a standard robust process model was required, the process needed to be a framework for carrying out the big data projects which would make the study less expensive and easy to replicate and easily manageable and thus CRISP-DM was selected based on the following advantages:

a) Independent of the industry sector
b) Independent of the technology used
c) Specifically designed for data mining projects

As mentioned by (Wirth, 2000) CRISP-DM is a comprehensive process model and is specifically for the data mining projects.

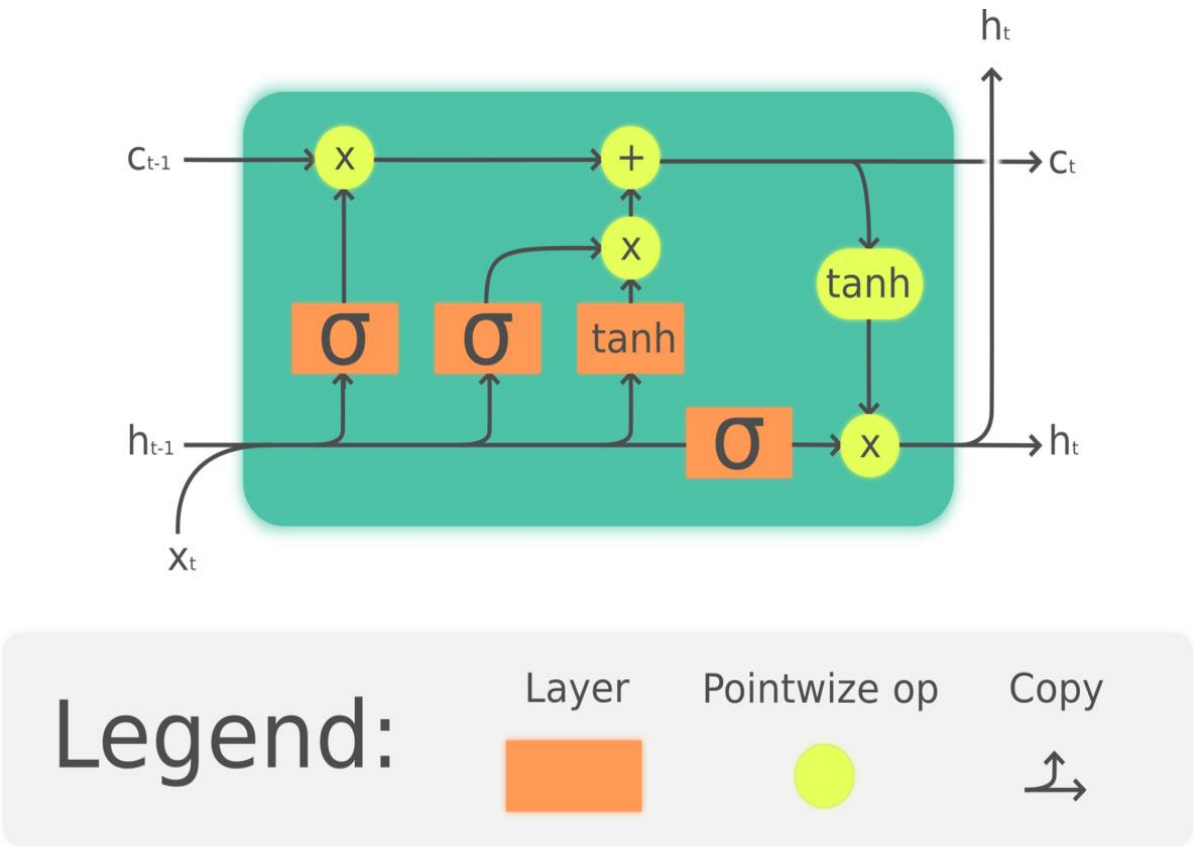


In this study the above CRISP-DM methodology is followed from understating the business requirements to implementation and evaluation and finally deployment of the model. The process was carried out as follows:

- ✓ The data fetched from smart meters was downloaded from http://www.ecodriver.uk.com/eCMS/viewfiles.asp?folder=DFID
- ✓ Downloaded data is imported in Jupyter Notebook (Python) as a data frame
- ✓ Only water meter readings are extracted into the data frame for the desired location
- ✓ The available data frame is pre-processed to remove any missing values and NA values.
- ✓ The pre-processed data frame is fed to LSTM model for prediction
- ✓ Here current reading is treated a X variable and next reading is treated as Y variable
- ✓ The predicted values are then evaluated against the actual values
- ✓ At every stage business question is reviewed to keep a check on the study
- ✓ Different Epochs were tested manually to run experiment again and again
- ✓ After a concrete model was ready with minimal RMSE, it was prepared for deployment.

Before finalizing the LSTM algorithm, many traditional and new algorithms were taken into consideration. Selection for algorithm was based on multiple factors as discussed in related works, each algorithm had its own pros and cons. Each algorithm brought in its own set of advantages and disadvantages. The selected algorithm was supposed to align with criteria

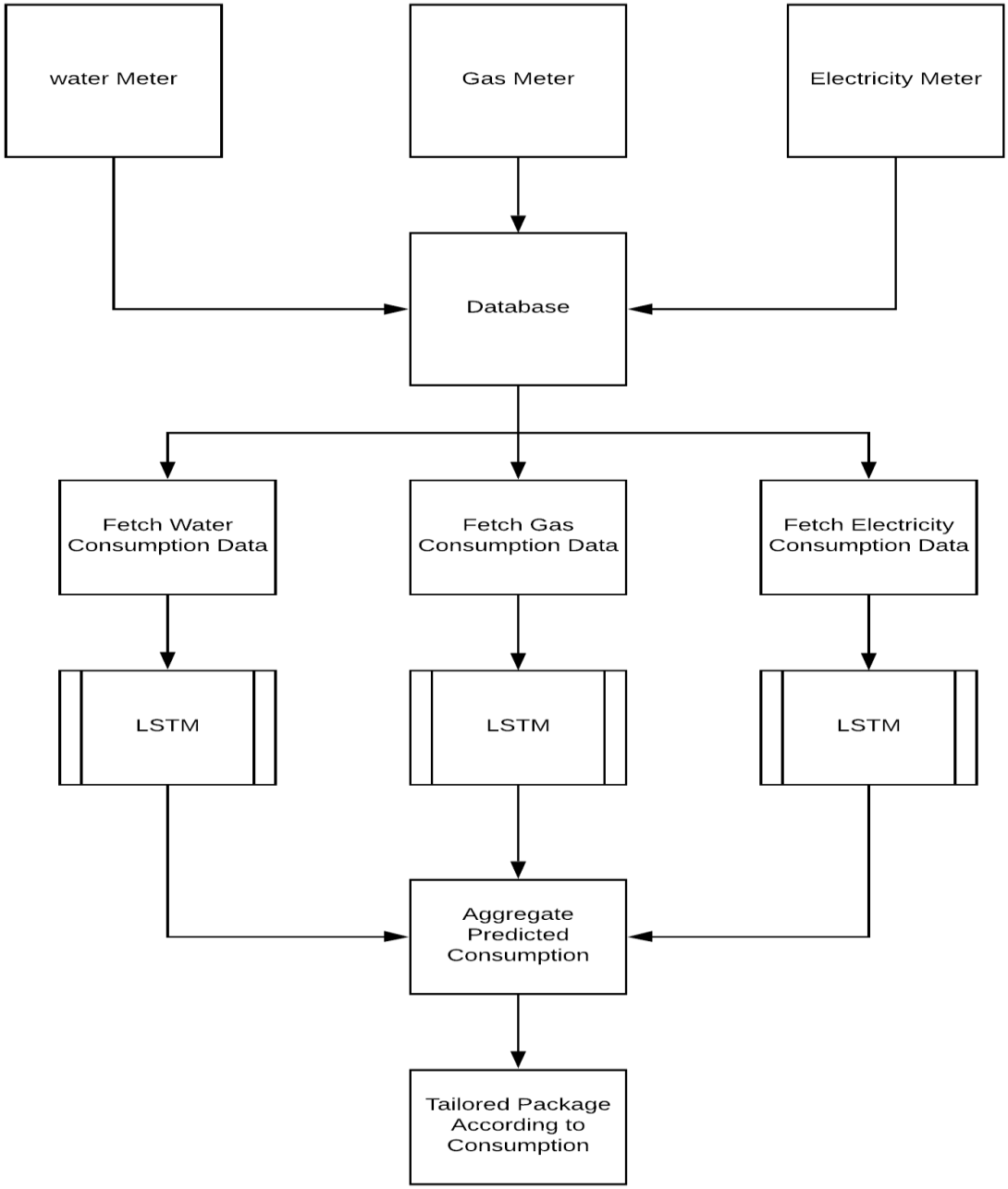
mentioned above for achieving the papers goal. The (Leyli-Abadi, Samé, *et al.*, 2018) which inspired this study adopted Joint Non-Homogeneous Hidden Markov Model is also a good model and provided 80% accuracy in (Leyli-Abadi, Same, *et al.*, 2018) research for forecasting the water consumption behaviour, although this model performed well as compared to other traditional models but as pointed out by (Computational and Unit, 2001) hidden Markov model requires initial probabilities to be defined by the user which should be randomly assigned. This allows the model to perform poorly if the assigned initial probabilities are not in line with desired values. This not only limits the prediction accuracy at a certain level but can perform poorly for time series prediction as compared to LSTM which was developed for time series forecasting in the first place, (Gers and Cummins, 1999) demonstrated how LSTM works very efficiently with the time series data and presence of outliers and missing values hardly impact the performance of the model. Unlike other forecasting model's LSTM has a forget gate that enables it to reset the learning process at appropriate interval and thus outperforms the traditional Recurrent Neural Networks. The reason behind RNN's disadvantage is the vanishing backpropagated error, and due to this RNN's lack the ability to learn the time lags greater than 5-10 discrete time steps.
Considering all the factors contributing in the LSTM model for time series forecasting, i.e minimal RMSE and ability to forecast with missing values present in the data set, LSTM was final choice for forecasting model using the historical data.



4 Design Specification

The data for this study is fetched from smart meters installed at each location and then aggregated in one database according to the location and time. After pre-processing the available data, the processed data is fed to the LSTM model for forecasting the next months consumption. Later the forecasted values are compared to the actual values and error rate is calculated to evaluate the model performance. When the error rate is minimal the model is deployed for forecasting.

The following flow-chart shows how the model is implemented to provide the tailored consumption packages to the consumers:

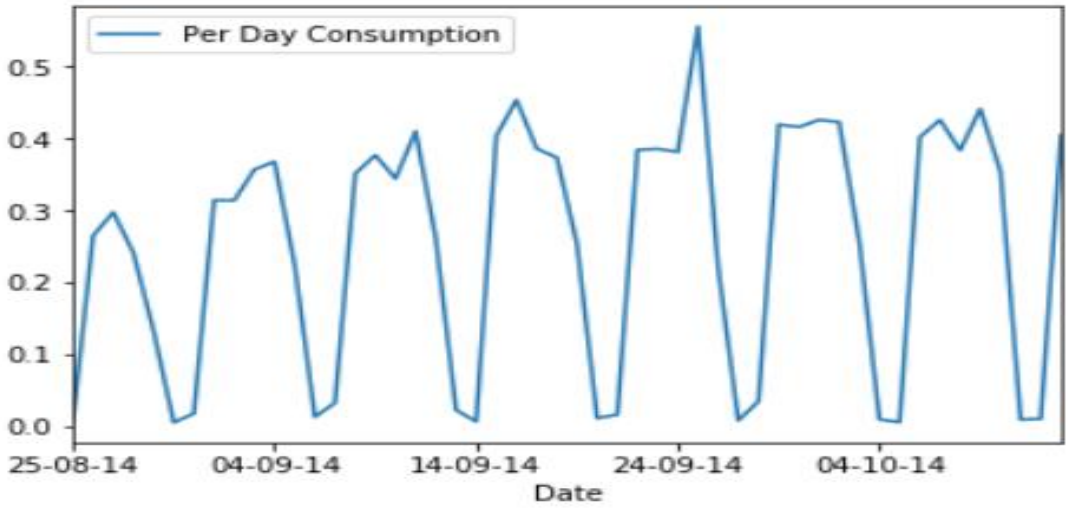


5 Implementation

Data Preparation

Even before implementing the forecasting LSTM model, it is necessary to prepare the data, the following transformations were performed on the data prior to the training and testing of forecasting:

a) Since this time series data had a increasing trend i.e increasing consumption need with passage of time, it was necessary to remove that trend and to make the time series stationary so as to have lag = 1.
b) Secondly to organise the data in a manner where past observation acts as an input to the model and current observation as the output generated by the model.
c) Transform the resulted data to have all the values between -1 and 1 for the activation function of LSTM model (hyperbolic tangent)
d) The transformed data is inverted on prediction so to scale them to original size for computing the error rate associated with the output.



Data Split

Following the traditional way of splitting the data, the consumption data was divided into two parts: 80% of the data the was kept for training whereas the remaining 20% was reserved for testing the model.

The data fetched from the source had smart meter readings from 2010-2019 for every 30 minutes, but total water consumption readings were released only in 2013 and the following years. Thus 5 years of consumption data is used for training purpose and the remaining two years data is kept aside for model testing. The forecasting model will be developed with the help of training data of five years and the final model will be tested for error with the help of testing dataset.

Experimental Runs

a) All the experimental scenarios were repeated ten times.

b) The need for these many runs is avoid the different results obtained due to random initial conditions assigned to the model.
c) For model configuration investigation a diagnostic approach will be used. The changes in the number of Epochs and then plotting the line charts will provide better insight of the model performs and how the parameters can be adjusted to improve the performance.
d) The designed model will be tested with both training and testing data so as to check if the model overfits or underfits with the help of RMSE
e) At the end of each experimental run the RMSE values will be printed for both training and testing dataset to observe the progress.
f) The training and the testing RMSE scores are plotted in the end to have a visual look for testing the model fitting.

Tuning Epoch Number

a) Epoch number is the first parameter that was changed at an interval for tuning the model during training time.
b) The Experiment started with the batch size of four and only one neuron, and different number of Epochs were tested to obtain minimum RMSE

Tune Activation Function

a) The activation and non-linearity of each neuron is controlled by the activation function.
b) There are many activation functions available in the Keras library e.g rectifier activation is the most popular activation function and few other activation functions such as sigmoid and tanh does a good job.
c) Despite so many activation functions available to us, we chose sigmoid activation function since we will be having only two classes for classification.

Tune Dropout Regularization

a) In order to limit overfitting of the model and increase the ability of the model to generalize, dropout rate is tuned. Here kept the dropout rate at 0.2 and gradually changed to 0.8
b) This process includes fitting weight constraint and dropout percentage.

Tune Neuron number in Hidden Layer

a) Number of neurons represent the capacity of network in that point of topology.
b) Here we started from 1 neuron and gradually increased till we get consistent result with minimal RMSE values.
c) The point where increasing number of neurons did not drastically affect the results was to be selected the ideal number of neurons.

5.1 Experiment 1/ Batch Size 1

LSTM Layers	Dropout	Epochs	Optimizer	RMSE

4	0.2	1000	Adam	0.53
4	0.3	1500	Adam	0.66
4	0.4	2500	Adam	0.45

5.2 Experiment 2/ Batch Size 5

LSTM Layers	Dropout	Epochs	Optimizer	RMSE
4	0.2	1000	Adam	0.15
4	0.3	1500	Adam	0.33
4	0.4	2500	Adam	0.35

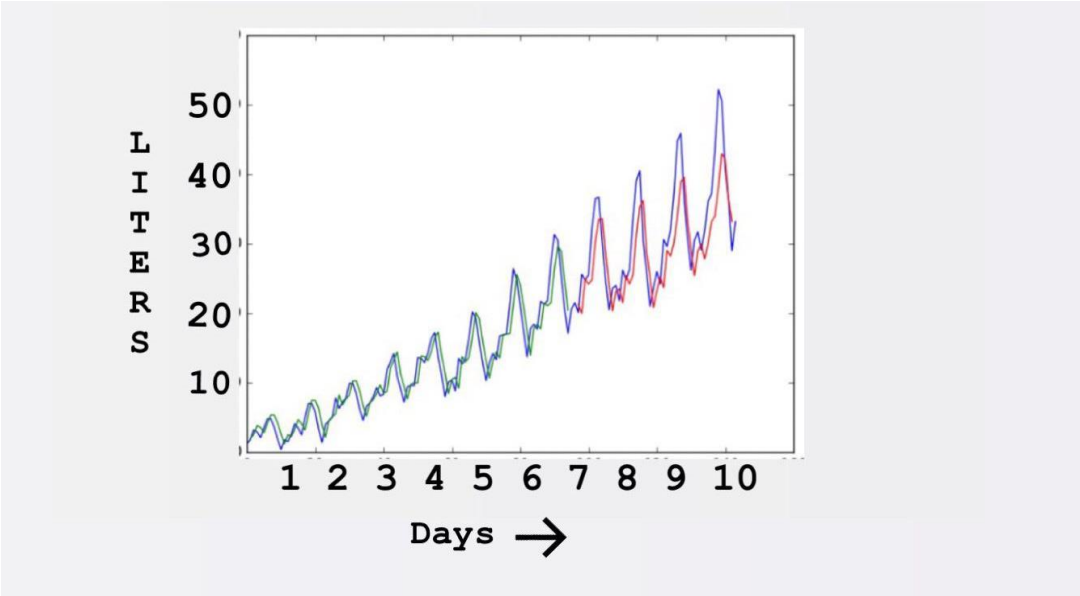
5.3 Experiment 3/ Batch Size 10

LSTM Layers	Dropout	Epochs	Optimizer	RMSE
4	0.2	1000	Adam	0.15
4	0.3	1500	Adam	0.16
4	0.4	2500	Adam	0.21

6 Evaluation

Unlike the evaluation criteria approach in (Leyli-Abadi, Same, *et al.*, 2018) of using accuracy as measure to check the performance of the model, here we took RMSE values for evaluating the model forecast. The reason behind this change in criteria is the study published by (Res, Willmott and Matsuura, 2005) that RMSE will give the idea of how far the predicted value is than the actual value. Whereas in accuracy the model performance seems to drop drastically if the numeric forecasted value is not the same as the actual value. So even if the model performs well and forecasted values are not exactly matching the actual values then accuracy will be poorly represented.

The following chart shows the forecasted data in comparison with the actual data:



As we observe from the above chart that LSTM was successful in minimising the RMSE values after each iteration and provided almost accurate consumption need for the following days. Here we also observe that training data has less RMSE as compared to the testing data, the reason for this deviation is overfitting of the model on the training dataset. This much overfitting is tolerable until we are getting a more precise forecasting for the testing dataset. Also, if we increase the number of Epochs and keep the dropout rate constant to 0.2, we can observe significant drop in the RMSE value irrespective of the batch-size. But after 2000 Epochs the RMSE value doesn't change much and increasing the layers would increase the computation power but no significant improvement in the performance is observed.

		LSTM(RMSE)
Water Dataset	Training	0.15
	Testing	0.21
Gas Dataset	Training	0.16
	Testing	0.20
Electricity Dataset	Training	0.16
	Testing	0.23

6.1 Discussion

Implementation of LSTM for time series forecasting works better as compared other prediction models discussed in the related works, although LSTM gives very low RMSE value than of other ANN's and probabilistic models it consumes a lot a computation power and since output of one neuron is fed to the next then like other models the process cannot be paralleled to i.e use all the cores available in the system to speed up the process. This not only put all the load on single core but increases the computation time, and as the dataset increases along with population and consumption needs the computation power and time increases drastically. Since the data is fetched at an 30min time interval for all the smart meters and stored in the database for all the years, this takes up a lot of space, specially when water, gas and electricity readings are stored in one location.

Although LSTM performs well for the forecasting, has very limited tuning options, the initial weights and values are randomly assigned and require several iterations for the experiment to have a stable consistent output. Also, LSTM can take multiple input features based on the dataset available, some studies discussed above included temperature and rainfall of the area to forecast the consumption behaviour but here only historical data of consumption is taken into the consideration.

For evaluation metrics many studies relied on the accuracy of the model, but here to evaluate the model RMSE was used. The reason behind not using accuracy as a metric was if the predicted value does not match with the actual value then accuracy will be zero even if the model performed well enough. For the following experiment different Epochs, batch sizes, optimizers and LSTM layers were tried in order to optimize the model, it was necessary to do this manually. Also, we need to invert the output of the model before we calculate the error score so that the performance of the model can be checked by comparing the output with the original values. Here LSTM helped in preventing the vanishing gradient problem which is commonly observed in RNN's but does not completely prevent it.

The Demand Side Unit implements similar ideas for forecasting the electricity consumption and provide them with night savers pack to save bills on electricity during their peak consumption, this study would contribute in adding water and gas consumption for similar profits.

7 Conclusion and Future Work

This study started with the goal to help cities and countries develop sustainably by carefully managing the resources such as water, gas and electricity. If the governing bodies could have the knowledge of how the demand for these resources is increasing and understand their future behaviour, then they can take appropriate measures for minimising wastage and promote optimum usage of these resources. The goal of this study was to forecast the consumption behaviour of water, gas and electricity for better management of these resources. The study assumed that consumption of these resources wasn't constant and had peaks and lows at different time periods and thus accurately forecasting the consumption is of paramount importance.

Here we implemented LSTM model for forecasting the consumption behaviour using historical data for next one month. The model performed successfully and gave RMSE value of 0.21 which was better than all existing models. It was found out that there were peaks and drops in the consumption behaviours and it changes (increases) in a specific manner. As compared to other models which were based of accuracy this model was evaluated based on RMSE and thus hold a

In future work, all these three consumption behaviours can be clubbed together to offer a tailored plan to the consumers thus reducing the prices and improving the working timetables for respective departments.

8 Acknowledgement

I would like to thank my guide Dr Anu Sahni who gave me this opportunity to work on this project , helped me in doing a thorough Research and I learned so many new things I am really thankful.

Also I would also like to thank my parents and friends who helped me in completing my research on time and encourage me to strive for the best.

References

A, D. S. & D. (2017) *International Water Scarcity and Variability : Managing Resource Use Across Political Boundaries*. California.

Adamowski, J. *et al.* (2012) 'Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada', *Water Resources Research*, 48(1), pp. 1–14. doi: 10.1029/2010WR009945.

Aksela, K. and Aksela, M. (2010) 'Demand Estimation with Automated Meter Reading in a Distribution Network', *Journal of Water Resources Planning and Management*, 137(5), pp. 456–467. doi: 10.1061/(asce)wr.1943-5452.0000131.

Chamroukhi, F. *et al.* (2011) 'Model-based clustering with Hidden Markov Model regression for time series with regime changes', *Proceedings of the International Joint Conference on Neural Networks*, pp. 2814–2821. doi: 10.1109/IJCNN.2011.6033590.

Computational, G. and Unit, N. (2001) 'AN INTRODUCTION TO HIDDEN MARKOV MODELS AND', 15(1), pp. 9–42.

Cutore, P. *et al.* (2008) 'Probabilistic prediction of urban water consumption using the SCEM-UA algorithm', *Urban Water Journal*, 5(2), pp. 125–132. doi: 10.1080/15730620701754434.

Dias, J. G. and Ramos, S. B. (2014) 'Dynamic clustering of energy markets: An extended hidden Markov approach', *Expert Systems with Applications*, 41(17), pp. 7722–7729. doi:

10.1016/j.eswa.2014.05.030.
Gagliardi, F. *et al.* (2017) 'A probabilistic short-termwater demand forecasting model based on the Markov chain', *Water (Switzerland)*, 9(7), pp. 7–14. doi: 10.3390/w9070507.
Gers, F. A. and Cummins, F. (1999) '1 Introduction 2 Standard LSTM', pp. 1–19.
Hokimoto, T. and Shimizu, K. (2014) 'A non-homogeneous hidden Markov model for predicting the distribution of sea surface elevation', *Journal of Applied Statistics*, 41(2), pp. 294–319. doi: 10.1080/02664763.2013.839634.
Kim, K. (2003) 'Financial time series forecasting using support vector machines', 55, pp. 307–319. doi: 10.1016/S0925-2312(03)00372-2.
Kwac, J., Flora, J. and Rajagopal, R. (2014) 'Household energy consumption segmentation using hourly data', *IEEE Transactions on Smart Grid*. IEEE, 5(1), pp. 420–430. doi: 10.1109/TSG.2013.2278477.
Labeeuw, W. and Deconinck, G. (2013) 'Residential electrical load model based on mixture model clustering and markov models', *IEEE Transactions on Industrial Informatics*. IEEE, 9(3), pp. 1561–1569. doi: 10.1109/TII.2013.2240309.
Leyli-Abadi, M., Same, A., *et al.* (2018) 'Mixture of Non-homogeneous Hidden Markov Models for Clustering and Prediction of Water Consumption Time Series', *Proceedings of the International Joint Conference on Neural Networks*, 2018-July. doi: 10.1109/IJCNN.2018.8489473.
Leyli-Abadi, M., Samé, A., *et al.* (2018) 'Predictive classification of water consumption time series using non-homogeneous markov models', *Proceedings - 2017 International Conference on Data Science and Advanced Analytics, DSAA 2017*, 2018-Janua, pp. 323–331. doi: 10.1109/DSAA.2017.32.
Liisberg, J. *et al.* (2016) 'Hidden Markov Models for indirect classification of occupant behaviour', *Sustainable Cities and Society*. Elsevier B.V., 27, pp. 83–98. doi: 10.1016/j.scs.2016.07.001.
Parks, R. and Mclaren, M. (2019) 'Briefing paper No 29 Experiences and lessons in managing water from Cape Town', (29).
Res, C., Willmott, C. J. and Matsuura, K. (2005) 'Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance', 30, pp. 79–82.
Şen, Z. and Altunkaynak, A. (2009) 'Fuzzy system modelling of drinking water consumption prediction', *Expert Systems with Applications*, 36(9), pp. 11745–11752. doi: 10.1016/j.eswa.2009.04.028.
Wei, S., Xingxing, Y. and Yanhong, L. (2008) 'The combination forecasting model for urban water consumption based on support vector machines', *Proceedings - 2nd 2008 International Symposium on Intelligent Information Technology Application Workshop, IITA 2008 Workshop*. IEEE, (3), pp. 805–808. doi: 10.1109/IITA.Workshops.2008.233.
Wirth, R. (2000) 'CRISP-DM : Towards a Standard Process Model for Data Mining', *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, (24959), pp. 29–39. doi: 10.1.1.198.5133.
Zhou, S. L. *et al.* (2000) 'Forecasting daily urban water demand: A case study of Melbourne', *Journal of Hydrology*, 236(3–4), pp. 153–164. doi: 10.1016/S0022-1694(00)00287-0.
Zucchini, W., Macdonald, I. L. and Langrock, R. (2017) 'Journal of Statistical Software', 80(August), pp. 1–4. doi: 10.18637/jss.v080.b01.