Literature review Imputation done by Ramon

What is this research for?	Summary or citation	APA
Variance error	The research of RAO and SITTER (1995) shows that the choice of variance estimators in imputed data differs largely based on several factors. A very important consideration that needs to be made is the susceptibleness to outliers for variance calculation.	RAO, J. N. K., & SITTER, R. R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, *82*(2), 599–631. https://doi.org/10.1093/biomet/82.2.453
K-algorithm	Zhang, Z (2016) stated that the K in KNN has a large impact on the diagnostic performance of KNN algorithm and that the key between an appropriate k value lies between a balance of overfitting and underfitting.	Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. *Annals of Translational Medicine*, *4*(11), 218. https://doi.org/10.21037/atm.2016.03.37
LOCF	Where the imputation methods KNN and LOCF will likely be sufficient for simpler cases of imputation. For larger chunks of missing data the former named methods were expected to have increasing error. However these methods should perform equally well for small percentages of missing data (Beck et al.., 2018)	Beck, M., Bokde, N., Asencio-Cortés, G., & Kulat, K. (2018). R Package imputeTestbench to Compare Imputation Methods for Univariate Time Series. *The R Journal*, *10*(1), 218. https://doi.org/10.32614/rj-2018-024
BMS or energy usage	A study that presents a computationally efficient data processing method for imputing missing data. This made use of a continuous profile with respect to adjacent available measurements, which is desirable for time-series analyses (Peppanen et al.,2016).	Peppanen, J., Xiaochen Zhang, Grijalva, S., & Reno, M. J. (2016). Handling bad or missing smart meter data through advanced data imputation. *2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. https://doi.org/10.1109/isgt.2016.7781213
Weighted KNN	Cost & Salzburg (1993) stated that a weighted instance makes it less likely for a new instance to appear near an exemplar as the exemplar weight increases. This in theory increases the accuracy when imputing missing data for most cases.	Cost, S., & Salzberg, S. (1993). A weighted nearest neighbor algorithm for learning with symbolic features. *Machine Learning*, *10*(1), 57–78. https://doi.org/10.1007/bf00993481
Recurrent Neural networks	Handling missing values in multivariate time series data using Recurrent Neural Networks is challenging. Using popular	Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent Neural Networks for Multivariate Time Series with Missing Values. Scientific Reports, 8(1). https://doi.org/10.1038/s41598-018-24271-9

	imputation methods leads to a time-consuming prediction procedure and may impair the prediction performance. To address the above issues, a GRU-based model could be used which captures the informative missingness by incorporating masking and time interval directly inside the GRU architecture. (Che et al., 2018)	
Missing data in energy building	This paper addresses the gap in the current study of missing data treatment. It presents a comparative study of eight methods for imputing missing values in building sensor data.	Pazhoohesh, M., Pourmirza, Z., & Walker, S. (2019). A Comparison of Methods for Missing Data Treatment in Building Sensor Data. 2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE). https://doi.org/10.1109/sege.2019.8859963
Sensor building management system data imputation	In this paper, a framework is developed to address these two gaps. First, a validation data generation module is developed based on pattern recognition to create a validation dataset to quantify the performance of data imputation methods. Second, a pool of data imputation methods is tested under the validation dataset to find an optimal single imputation method for each sensor, which is termed as an ensemble method.	Zhang, L. (2020). A Pattern-Recognition-Based Ensemble Data Imputation Framework for Sensors from Building Energy Systems. Sensors, 20(20), 5947. https://doi.org/10.3390/s20205947
Bi-directional data imputation	This study proposed a methodology, namely the hybrid Long Short Term Memory model with Bi-directional Imputation and Transfer Learning (LSTM-BIT). It integrates the powerful modeling ability of deep learning networks and flexible transferability of transfer learning	Ma, J., Cheng, J. C., Jiang, F., Chen, W., Wang, M., & Zhai, C. (2020). A bi-directional missing data imputation scheme based on LSTM and transfer learning for building energy data. Energy and Buildings, 216, 109941. https://doi.org/10.1016/j.enbuild.2020.109941
Machine learning for energy data	this research aims to address the problems by developing a predictive model for energy consumption in Microsoft Azure cloud-based machine learning platform. Three methodologies which are Support Vector Machine, Artificial Neural Network, and k-Nearest Neighbour are proposed for the algorithm of the predictive model.	Shapi, M. K. M., Ramli, N. A., & Awalin, L. J. (2021). Energy consumption prediction by using machine learning for smart building: Case study in Malaysia. Developments in the Built Environment, 5, 100037. https://doi.org/10.1016/j.dibe.2020.100037