Received: 9 August 2023 | Revised: 20 November 2023 | Accepted: 27 December 2023 | Published online: 27 December 2023

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2025, Vol. 3(3) 231–239
DOI: 10.47852/bonviewJDSIS32021517

BON VIEW PUBLISHING

Data Science and Machine Learning Processes for IoT-Based Pulsed Plasma Thruster Research

Noman Hossain[1,2,*], Guorui Sun[3] and Farjana Islam[4]

[1]*Department of Computer Science and Engineering, University of Liberal Arts Bangladesh, Bangladesh*

[2]*Center for Innovation Technology, University of Liberal Arts Bangladesh, Bangladesh*

[3]*Tianjin Institute of Power Sources, China*

[4]*Bab Africa Industrial Engineering Solutions for Companies, Bangladesh*

Abstract: Pulsed plasma thrusters (PPTs) have very high specific impulse but low efficient electric propulsion engine, which are being used as primary or secondary propulsion mechanism for spacecrafts (cube, micro, nano, or pico satellites). PPTs are used in academic and/or industrial research from decades, but its experimental data collection methods are normally offline, analysis methods are conventional, and data interpretation lies in electromagnetic, physical, and chemical domain. It is inadequate to explain PPT-generated data in few specific domains only, and it may prone to inadequate explanations, which can hide vast insights of data. Actually, PPT-generated data are usually big data. It is essential to analyze and explain PPT data in data science and machine learning (ML) domain to get keen insights from the data. To this date, no such complete solution exists for PPT either in industrial arena or academia to use all these three technologies. To meet the gap, we propose and implement an Internet of Things (IoT)-based architecture for PPT experimental facility to collect, accumulate, and process PPT data intelligently. This architecture yields some ML models with prediction accuracy above 93%. This architecture is a complete IoT and big data-based artificial intelligence implementation on PPT data through data science and ML processes. This architecture is capable to contribute in both academic and industrial research, development, and deployment.

Keywords: artificial intelligence, big data, data science, electric space propulsion, Internet of Things, pulsed plasma thruster

1. Introduction

Pulsed plasma thrusters (PPTs) [1] are the source of huge data from experimental condition to onboard space condition as satellite module. PPT data are collected via a number of sensors or probs like current, voltage, magnetic field, temperature, flow, etc. Common practice of gathering these data is offline segregated manner. After collecting data, processing also requires a subtle amount of time, experience, and effort to analyze them and produce results. Difficulty becomes worst while a number of programming or technical skills are essentially required. After gathering the data, researchers use some data processing software to pull results from the experimental [2, 3] or simulation [4–6] data and try explaining the results in physical [7, 8] or chemical [9, 10] science domain. It is a way of explanation but not optimal solution because explain bare data with different theories other than data science domain, it may dig some hidden hole in which can hide actual meaning of physical world. So, it is innate demand that data science process should employ beside physical science explanation to explain physical processes of PPT more robustly. Till now, there is no noticeable endeavor regarding available PPT data explanation using data science and machine learning (ML) approach.

On the other hand, performing PPT experiments is costly in respect of time, money, and effort. It costs huge while repetitive occurrence of same experiments is a common phenomenon. So, there is a great chance of data redundancy and idleness of costly data. Internet of Things (IoT) [11] based approach can play a role to automatically accumulation of data in storage; thus, make a historical data footprint for future use cases and reuse them when needed and it also confirms that no previous experiments are done repetitively. This method takes very latency in time as compared to the previous human-driven manual data collection and execution of experiments.

Moreover, data processing in trivial manner requires one or more software, programming skills in some extend, and result validation. Sometimes we need to buy some commercial software, which is costly in nature and has a number of drawbacks in usage like software training, periodic maintenance cost, and dependency on others. In such a case, an open-source solution would be a great fit for academic or individual researchers. It would be more convenient and time efficient if there is any method available to get self-validated result by putting minimal or no programming effort. Here comes the ML solutions to play a role, which requires very low to no programming effort and skill to get quality results from on board, experimental, or

*Corresponding author: Noman Hossain, Department of Computer Science and Engineering and Center for Innovation Technology, University of Liberal Arts Bangladesh, Bangladesh. Email: noman.hossain1@ulab.edu.bd

simulation data. To achieve cost and effort effective, automated aggregation and processing, less redundant and easily deployable solutions for PPT data, we adopt a number of open-source technologies.

The following contributions are made throughout this paper:

1) We propose an IoT-based architecture for PPT data collection and analysis by using core data science and ML methods. We name it the tower-shaped data science and ML (TDM) architecture. It is the first endeavor of this arena.
2) Our proposed solution is capable to process and predict data either coming from onboard satellite or experiment or simulation. We perform our test run on PPT experimental data and get successful output as expected.
3) This architecture is capable to replace some software, which is used to gather and process data. It is also capable to act as a substitute to real experiments.
4) This architecture yields data model and services. User can use this data model to analyze their data and can use restful application programming interface (REST API) to get data from our system as a service.
5) User can use the output model of this architecture to make web or mobile applications, thus making these solutions more available and public to research community.
6) This TDM architecture is a feasible solution to minimize data redundancy, production cost, and processing effort.

The rest of this paper is organized as follows.

In Section 2, we describe the related work to seek the way of conducting our research. Section 3 explains our proposed TDM architecture in detail by each component. Section 4 describes the experimental setup for data science and ML processes, among them IV-A describes illustrative view of data science outputs and IV-B explains the illustrative view of ML processes. We describe results and respective discussion of TDM architecture in Section 5. Finally, we conclude the paper and outline future work to improve it in Section 6.

2. Literature Review

In electric propulsion research, PPT plays significant achievement in the development of recent cube satellite [12] revolution. This revolution demands involvement of multidisciplinary research to employ further revolution in upcoming era. Till now, PPT data are explained by different physical or chemical theories. Experimental procedures are normal data gathering and then process using some processing software.

IoT and data science methods are the mostly used data gathering and analysis tools by different [13–15] diverse fields. They are taking advantages of using these cutting age technologies. IoT-based solutions are reliable and effective in real-time and high-speed data transmission. In aerospace propulsion, involvement of neural network [16, 17], a form of ML, is practiced in some instances like rocket or aircraft use cases other than PPT domain. Only a few instances [16, 18, 19] are available in PPT research, which involve neural network to analyze PPT data for anomaly detection, but none of them have complete solution for real-life use cases.

We found following recent research articles where researcher tried to implement ML and other artificial intelligence (AI)-based methodologies for the PPT or close cousin of the PPT systems like Hall effect thrusters [20] data processing. However, there is a lot of scope to introduce a complete data processing architecture. Also, a detailed industry standard work is need to be introduced.

We found no significant and complete research is done recently in the PPT in experimental or real time or simulation domain using the three technologies used in the TDM architecture, though it is the demand of time. Primitive data processing and analyzing method are not cost effective and quite an old way of data handling.

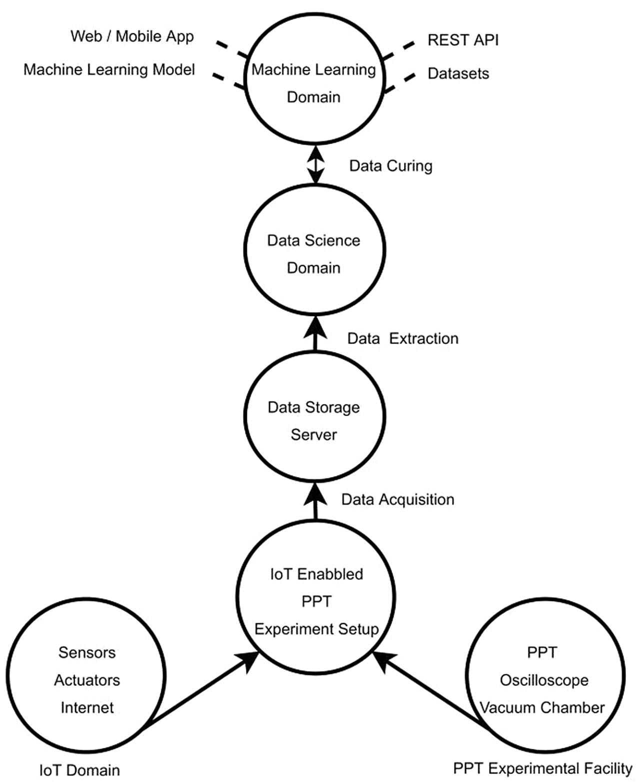
Recent innovation of modern technology especially AI, data science, and IoT technology is leading option in processing data. Processing and analyzing PPT data demand involvement of these technologies to keep continuing its development further. So, we intended to use IoT-based solution for PPT data analysis and prediction.

3. IoT-Based TDM Architecture for PPT

3.1. System design components

This study has employed a cross-platform system design in order to analyze the PPT propulsion by incorporating IoT, data science, and ML. The proposed TDM architecture for PPT experimental or simulated data is shown in Figure 1. This architecture is feasible to implement not only PPT experiments but also other experimental use cases either in native aerospace engineering or in different domain. Employing IoT technology to PPT experiments yields a number of advantages from data collection to data organization. While common methods are offline and human-driven solution in data collection and processing which demands enormous human effort and time. For better results and suitable explanation, a more efficient method of data collection and analysis was in the heart of the interest from long ago. The proposed TDM architecture can fulfill such a demand in large extent. TDM is an automatic solution with minimal human intervention.

Figure 1
The TDM architecture



3.2. The TDM architecture

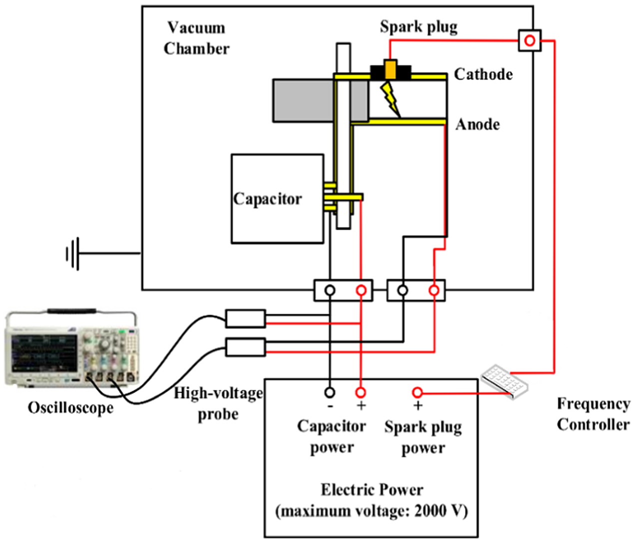
The proposed TDM architecture combines the existing PPT experiments setup with IoT realm, thus creating IoT-based PPT experiment facility. We name it TDM because the internal processing of data run from root of the tower to roof of the tower as shown in Figure 1. The data flowing from root are raw data generated by PPT experiment and data available in roof are processed final output pulled by our architecture. The data flow of TDM begins from the bottom of the tower to the roof of the tower. This architecture confirms healthy data flow through a number of scheduling and synchronizing scripts written for Linux system.

Actually, TDM architecture is a semi real-time architecture in data collection and processing method. Real-time data ingestion and processing is our future scope of work. In TDM, we minimally confirm the data flow for the maiden ML platform for PPT experiment. Rest of the TDM architecture is given in the following sections.

1) PPT experimental facility: The normal PPT experiments setup shown in Figure 2(a) [21] consists of three essential components, i.e., vacuum system for creating space environment, measurement system includes oscilloscope, sensors, actuators, and measurement probs to measure and gather data signals, and propulsion system. Space system includes the cylindrical vacuum chamber (0.5 m × 0.8 m), and it maintains vacuum environment by its two rotary vane pumps and one molecular pump. The propulsion system includes a prototype of ablative pulsed plasma thruster (APPT) shown in Figure 2(b) [21]. The measurement system consists of two voltage probs and one oscilloscope. We use built-in oscilloscope sensor and prob signal to log data into local computer.
2) IoT domain: In TDM architecture to create IoT facility, we reuse built-in oscilloscope sensor and voltage and current probs as the IoT sensor, and connect local computers with internet to process data up into the big data server storage. We use periodic batch processing of data during experiments.
3) IoT-enabled PPT experiment setup: To make IoT-based experiment setup for PPT, we rearrange the previous setup, recuse built-in sensors and previously available probs, interconnect the local computers with internet, and introduce a high-end big data server and some synchronizing and data transfer schedule for the batch uploading and processing of data. The IoT-based PPT setup is shown in Figure 3. This newly formed IoT setup works well and capable to transfer data to the preestablished big data server.
4) Data acquisition: Data acquisition (DAS/DAQ) is the process of examining sensor state or signals that measure actual experimental or physical conditions over time and changing them into computerized numeric qualities that can be controlled by a PC. In our TDM architecture, it deals with PPT experimental data stored in local system via current and voltage prob signals and processed by built-in sensors of oscilloscope and transfer them up into storage server in periodic batch processing manner. We can use tool like Node-RED, MQTT etc., To keep TDM operation simple and integrated, we write data acquisition scheduling script by ourselves. Initially, we use Linux Crontab scheduling mechanism.
5) Data storage server: Data are generated by the PPT discharge and captured by the oscilloscope sensor signals, which are published in csv format and saved in local computer. We can use different data storage servers like OpenStack or similar one, but for the simplicity of implementation to intended target users, we choose server version of MySQL database as a maiden test run storage. We use Apache Sqoop to import and export data from server to local system. This data storage server will act as a historical data footprint for PPT experiments. The naming convention to store PPT experimental data into storage server is "timestamp-PPT_type-experiment_type-experiment_number," e.g., 2019-09-12-les6-life_time_testing-exp1. To ensure data integrity and naming convention, we developed a script to preprocess the data before saving into storage server.
6) Data extraction: The data extraction process is responsible to extract, export, and download quality data from server to data science and ML environment with full integrity. We choose Apache Sqoop to export data from server to local machine

Figure 2
Non-IoT PPT experimental facility

(a) **Normal PPT experiment setup**



(b) **APPT**

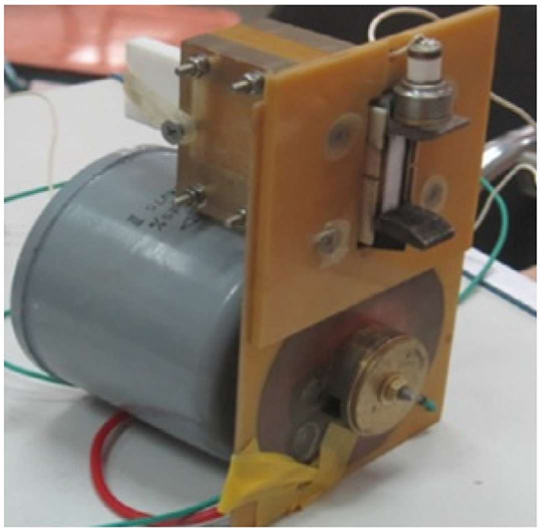
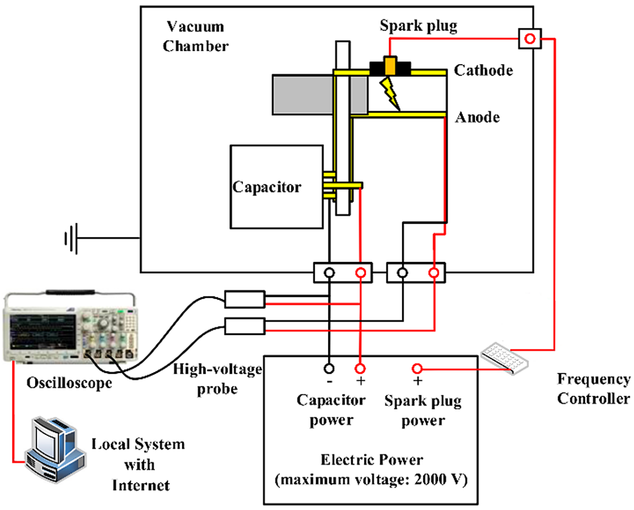


Figure 3
IoT-enabled PPT experimental facility



where environment is set up for further processing of data. Besides this, we can use direct URL to locate data into local environment from server. The extraction commands follow a Crontab scheduling script or it can run on demand.

7) Data science domain: Data science is a vast domain to understand, cure, analyze, and process data. It is also an umbrella term, which encompass multidisciplinary sciences like statistics, mathematics, and computer science amalgamate together. The normal data science processes have the following parts shown in Figure 3. As data science covers multidisciplinary sciences together, some of the parts of data science is common to others science. For example, data modeling is the constituent part of both data science and ML. In this paper, we treat model building as the parts for ML and we use data science for data preparation and exploration only. Our main intention regarding data science process is to prepare data and explore data to understand them. We use Python ecosystem for data science processes.
8) Data curing: Data curing process deals with health of the data. It is done by data preparation step of data science processes shown in Figure 3. Data curing process ensures no faulty data are going to ML processes in the next stage. Actually, most of the steps of data science processes do data curing to make data fresh and faultless. Initially, there are 10,000 rows in our PPT raw data. After curing process, we get a fresh and workable dataset having 2800 rows. Data curing process reduces the redundancy of data and ensures quality.
9) ML domain: ML is the alternative way of data processing by which a computer can work more accurately as it collects and learns from the data given. It is one of the state-of-the-art techniques to process data. It is one of the parts of data science processes, but we treat it differently as a separate domain. There are a number of ML tools available in open source like Apache Mahout, Apache Spark, PyTorch, TensorFlow, scikit-learn, and many more. We choose Python scikit-learn as our ML platform for its simplicity and robustness. In ML processes, learning algorithms play the key role to train and test data. It is possible to develop an algorithm by ourselves but not a wise idea. Most of the ML platforms have a set of optimized algorithms. We just need to select the right one from them and use it. We primarily test a set of linear and nonlinear algorithm such as linear regression (LR), random forest (RF), support vector machine, gradient boosting method (GBM), and many more. We found GBR algorithm has the best accuracy against our dataset.
10) Datasets: Datasets are the final and refined output of raw data processed by data science process. ML processes are mostly depending on the health of data. If data are healthy, then outputs from ML process demand value in real life. Our final dataset contains 3 features, i.e., time in (μs), voltage in (V), and discharge current in (A). It is the first direct offspring of the work.
11) ML models: ML models are the key factor for leaning from data. Only suitably trained models can elicit right pattern from datasets. It is crucial to get right data model from data because faulty model can produce anomaly prediction or classification. In our case, we successfully build a model with accuracy of 98.999%. We use root mean square error (RMSE), root square (R2), standard deviation, and variance to test our model accuracy. This model is another offspring of the work.
12) REST API: REST API is the heart of the data service provided for web or mobile application. We develop a REST API to ensure data service for researcher. Interested parties can use this REST API to request and generate data for their need. To generate same quality data as like as experimental data one can, use this API to build new datasets for PPT. Quality of data will be same as compared to real experimental data. In future, after performing some test, this REST service can act as a substitute of real costly experiments. It is the more public offspring of the work.
13) Web/mobile app: One can develop web or mobile application either for PPT or any other IoT-based application by using our REST API. It is the indirect offspring of the work.

4. Experiment Executions

To perform data science and ML process on PPT data, we present an experiment setup assisted by IoT. We use two computers, among them one machine acts as the big data platform as well as data processing platform and another machine acts as the data logger role through IoT networks. Server computer is a state-of-the-art computer having Intel 8th generation core i7 processor, consisting 12 CPU core inside. We use CentOS 7.4 as an operating system and Spark as the big data platform. We also completed IoT-based PPT experimental setup beforehand as described in the previous section. The description of the detail data processing is given in two sections as follows.

1) Data science processes: Data science process starts from first loading the data located either from server or downloaded into local storage. The overall data science process is shown in Figure 4. It starts from setting the goal of the process to model the data. We meant data science process is the process before data modeling. Data modeling part will be covered in the ML sections. We use Python Pandas package to load and read csv data of our PPT dataset. After loading required packages and data, we start checking the internal structure of data by descriptive statistics like shape, types, peak, and summery estimation. These are helpful to understand the anatomy of data. Then, we perform univariate data analysis of each variable to confirm their distributions and mean, median, variance, and density. The histogram plot of PPT data is shown in Figure 5; it is clear that discharge current and

voltage are instable and have skew in distribution. We also observe that density distribution of voltage has abnormal state, which is shown in Figure 5. From bar plot in Figure 6, it is observed that time has exact mean, voltage has mean toward quarter percentiles, and current has a lot of outliers. We can some up that voltage and time are smoother in distributions, but discharge current is a bit unstable. In reality, discharge current is unstable in PPT experiments due to plasma instability.

Figure 4
Data science processes

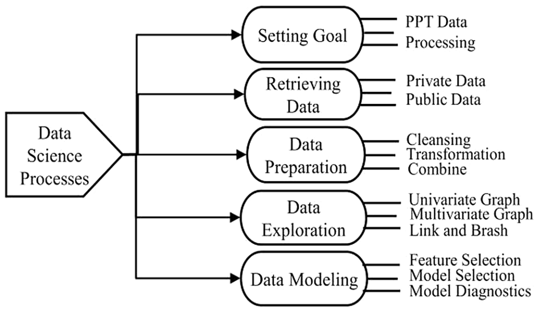


Figure 5
Bar plot of each PPT variables

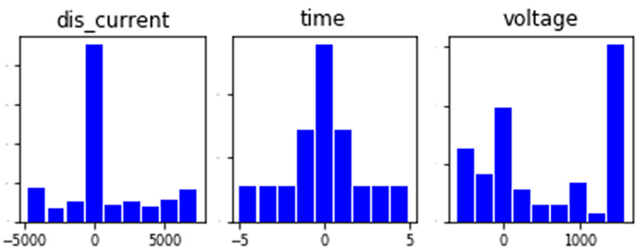
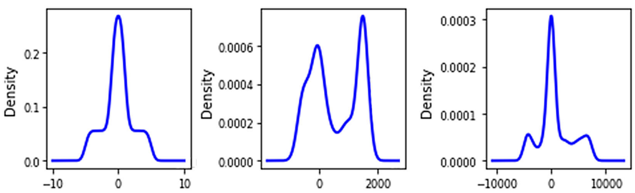


Figure 6
Density plot of each PPT variables



Afterwards, we perform multivariate analyses to find relationships among variables. Correlation and matrix plot are usually used for multivariate analysis. By norm, mostly correlated data are the source of wrong prediction, so they are the target for removal from datasets. Figures 7 and 8 visualize the correlations between PPT variables. We see no such correlation between PPT variables to remove it. Current and voltage variable pose spiral relationship between them.

2) ML processes: We start our ML process in finding a best algorithm to train our data and finally build a model to predict new data from none. The overall data modeling and training process is guided through the steps given in Figure 9. The algorithm selection is vital in this step. Output and accuracy of learning model are mostly depending on algorithm. Another thing is that, actually, we have no idea which algorithm will do best for our data in prior to test them. We test a number of linear and nonlinear algorithms including LR, Lasso regression and ElasticNet, classification and regression trees, support vector regression and *k*-nearest neighbors, AdaBoost (AB), GBM, RFs, and extra trees (ET), and choose GBM algorithm for our dataset to process it further. The statistics is given Table 1, and result is shown in Figure 10. Next, we start building our model using GBM algorithm. After building it, we train and test them accordingly. We also calculate RMSE, R2 score, and variance.

Figure 7
Box plot of each PPT variables

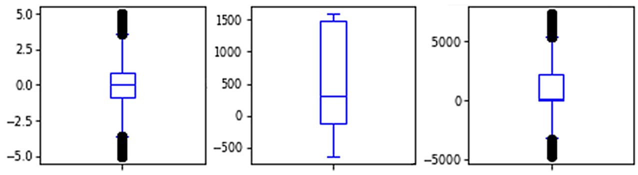


Figure 8
Scatter plot matrix of each PPT variables

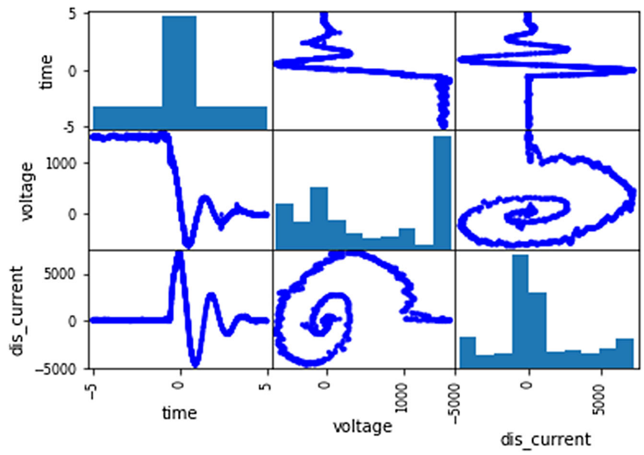


Table 1
Algorithm comparison

Algorithm	Mean	Standard deviation
Scaled AB	−7017.08	737.44
Scaled GBM	−616.94	108.40
Scaled RF	−436.39	74.92
Scaled ET	−430.57	81.64

The R2 score for both test and train is 0.9998. RMSE for overall case is 486.20 and variance is 1. It is very important that variance should be 1 or near to 1. This is the end of the ML process and we will discuss our outputs in the results and discussion section.

Figure 9
Data modeling process using ML algorithms

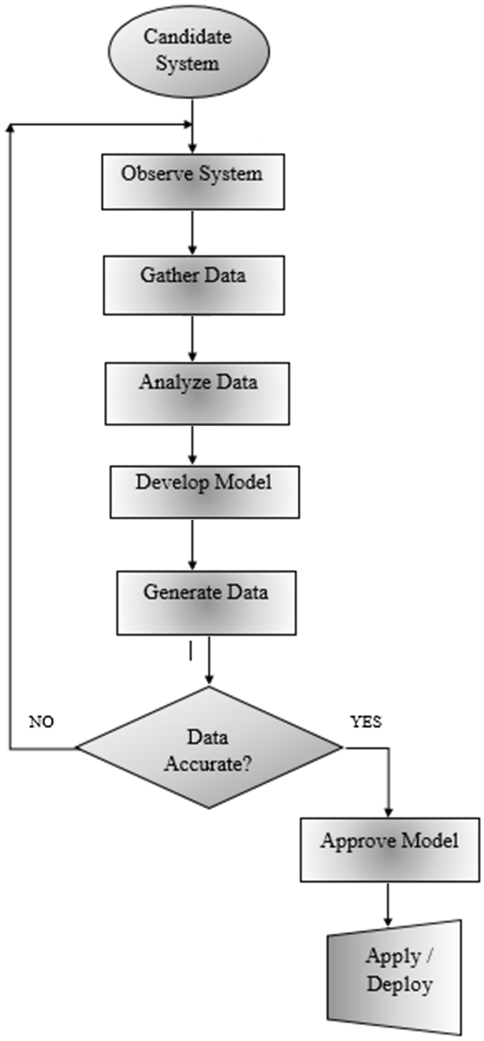
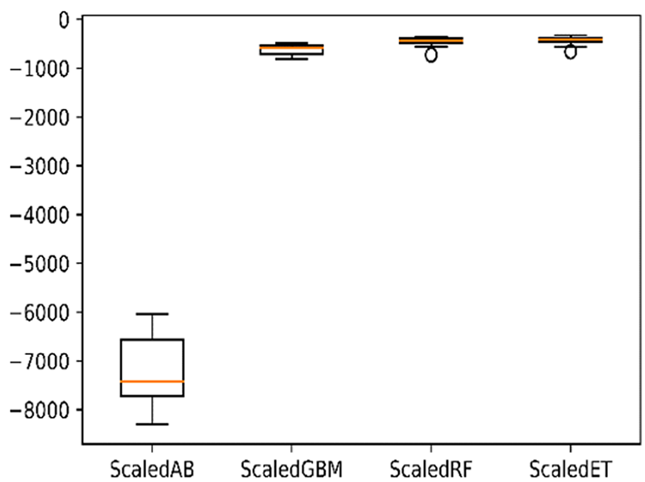


Figure 10
Comparisons of ML algorithms

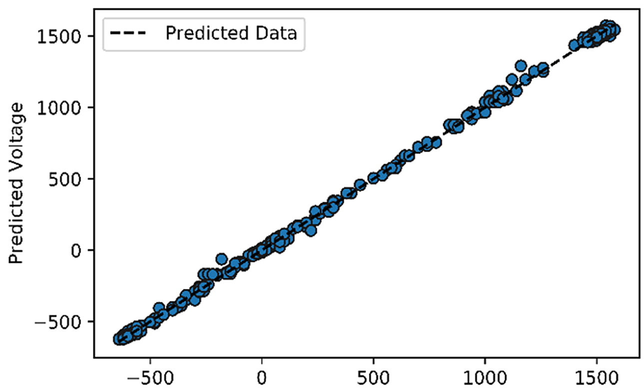


5. Results and Discussion

ML process yields four outputs. Two of them have noticeable impact on experimental process of PPT. At first, we will discuss the robustness of our developed model.

The selected GBM algorithm trains the model in such a manner that the learning outcome for the case of voltage is fully converged with real-life experimental data as shown in Figure 11. The input feature to training model is time (μs), and it can generate empirical set of data.

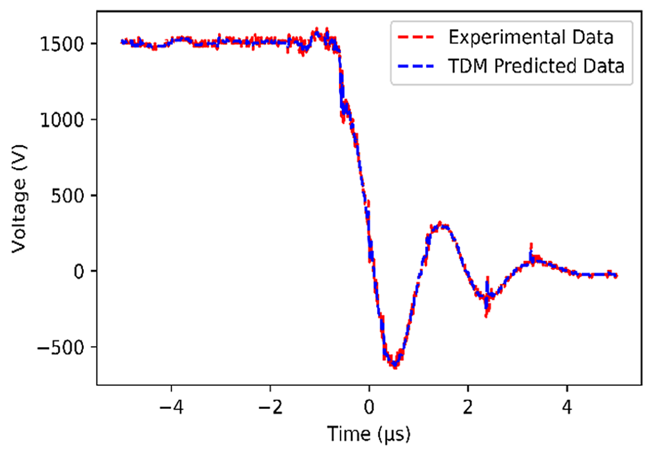
Figure 11
Ground truth vs predicted voltage



We have tested these results in our developed RESI API; it also yields same set of output voltages. It is clearly visible that predicted data fit with actual data having minimal error.

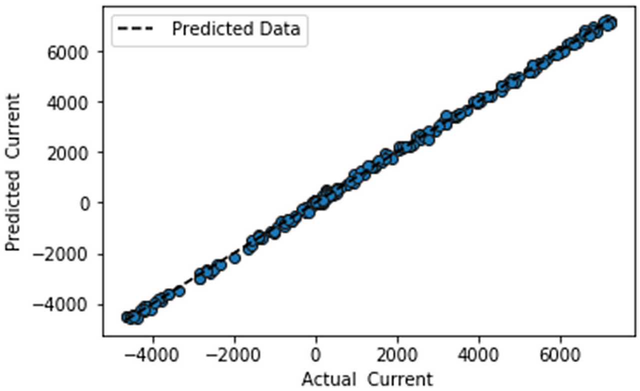
To clarify more, we draw another sketch of time vs voltage graph shown in Figure 12. The red dashed line belongs to experimental data and blue dashed line is the predicted results by our ML model, TDM. The result is so smooth that it neither overfits nor underfits ever. This smoothness of output graph implies the robustness of this model.

Figure 12
Experimental vs predicted voltage



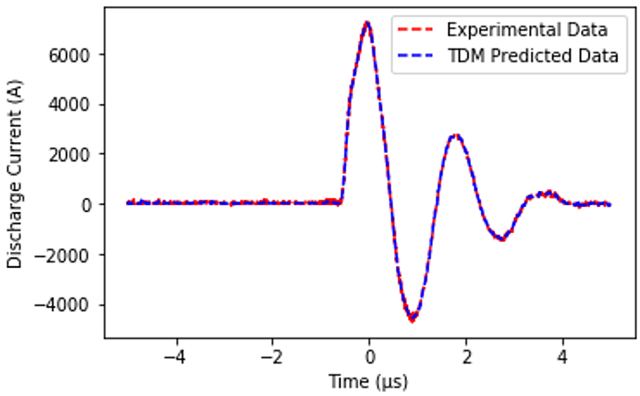
The optimized GBM algorithm trains the model in such a manner that the learning outcome for the case of current is also fully converged with real-life experimental data as shown in Figure 13. The input feature to training model is time (μs), and it can generate empirical set of data.

Figure 13
Ground truth vs predicted discharge current



The test results for the case of discharge current in our developed RESI API are promising; it also predicts same set of output discharge current as done in real experiment. To clarify more, we draw another sketch of time vs discharge current graph shown in Figure 14. The red dashed line belongs to experimental data and blue dashed line is the predicted results by our ML model, TDM.

Figure 14
Experimental vs predicted discharge current



Finally, we will discuss about the outputs that will impact real-life experimental situation. The best deliverable outcome of TDM architecture is the ML model, namely "ppt-tdm-1500V-voltage-model" and "ppt-tdm-1500V-current-model". These models can predict the voltage and current values of LES6 PPT, respectively. These models can be used to develop web or mobile applications to reach more usability for researchers.

We develop a REST API and use our ML model to provide data service to intended community. This REST API is constantly pinging exact solution compared to real experimental data. To get output from our RSET API just hit following commands from your Linux terminal for the voltage prediction.

curl -X POST http://localhost:5000/predict_voltage -H 'Content-Type: application/json' -d '{"features": [1.8]}'

The terminal will respond the following as the predicted output.

{"prediction": 155.1777834}.

The above REST command sends request to the API to predict the voltage of time 1.8 μs. API will response to it and sends back the 155.17 V as the predicted result. This result is 160 V in the experimental case. So, our result 155/160 is a best prediction. This also signifies our contributions in this regard.

Some randomly chosen voltage response of REST API is shown in Table 2 with error rate. The data output of the above table signifies the low error rate of prediction as claimed before. This outputs also evaluate our contributions.

Table 2
TDM API performance on voltage data

Input time (μs)	Real data (V)	TDM predicted data	Efficiency in prediction
1.8	160	155	155/160 = 0.93
1.82	120	119.8	119.8/120 = 0.99
1.89	40	40.8	40.8/40 = 1.0
2.0	−20	−6.7	−6.7/−20 = 0.3
1.8	160	155	155/160 = 0.93

Similarly, to get discharge current output from our RSET API just hit following commands from your Linux terminal.

curl -X POST http://localhost:5000/predict_current -H 'Content-Type: application/json' -d '{"features": [1.8]}'

The terminal will respond the following as the predicted output.

{"prediction": 2667.28841974}.

The above REST command sends request to the API to predict the discharge current of time 1.8 μs. API will response to it and sends back the 2667.28841974 as the predicted result. This result is 2666 A in the experimental case. So, our result 2667.28/2666 is a best prediction. This also signifies our contributions in this regard.

Table 3
TDM API performance on current data

Input time (μs)	Real data (A)	TDM predicted data	Efficiency in prediction
1.8	160	155	155/160 = 0.93
1.82	120	119.8	119.8/120 = 0.99
1.89	40	40.8	40.8/40 = 1.0
2.0	−20	−6.7	−6.7/−20 = 0.3
1.8	160	155	155/160 = 0.93

Few randomly chosen current responses of REST API are shown in Table 3 with error rate. The data output of the above table signifies the low error rate of prediction as claimed before. This outputs also evaluate our contributions.

The best feature of TDM architecture is that it can generate experimental quality of data based on input variables. So, this TDM architecture can be acted as a substitute of real costly

Table 4
Cost comparison

Parameter	Experimental	TDM
Mode of data generation	Operational	Prediction
Time cost	6–12 h	Few seconds
Human effort	Intensive	Minimal
Financial cost	Substantial	Minimal
Human role	Control, process, analysis	Provide input features

experiments. The cost comparison of TDM architecture is given in Table 4.

From Table 4, we can see a number of parameters that entail real experiments to generate PPT data and after generation they have to process to pull output by using different methods and software. The costliest parameter is time cost, because it takes 6–12 h to complete experiments to get data from PPT, whereas TDM architecture can generate same data within few seconds. A similar work used similar approach is be best mention in Hossain et al. [18], where authors developed a data-driven model to asses PPT behavior using AI.

Another issue is the financial cost; it is very costly to run vacuum systems of PPT experiments to create very low-pressure space environment vacuum condition, whereas we can minimize this cost to zero by using TDM API. Other cost are variables in experiment cases like human effort and role, but it can be made ensure to minimal using our method.

6. Conclusion and Future Work

In this paper, we proposed TDM architecture for processing and predicting PPT experimental data using data science and ML methods. We develop TDM architecture using open-source components to make it useable for research community. We achieved prediction accuracy of ML models above 93% using this architecture. We develop a REST API for data service to pull new data without doing experiments. We tested the performance of TDM architecture in different ways and found our proposed model is effective in data prediction. And it ensures human interventions and involvements to minimal. So, TDM architecture can be used as a cost-effective solution to PPT research.

Although our TDM architecture is working well and yields best efficiency, a number of things can be done to make it more adaptive. We did only periodic batch data collection using simple IoT method. In future, we will do real-time analysis of PPT data and predict the PPT performance and health diagnostics in real-time. We will do more to increase robustness of TDM data ingestion and processing using big data platform and analytics.

Recommendations

The implementation of AI, big data, and IoT provides ease of data modeling and processing through TDM architecture. Therefore, utilization of the TDM approach is recommended for both academic and industrial space propulsion research. Since less effort and cost is required to process PPT data using TDM architecture, extended use of it also recommended beside the traditional simulation and experimental data processing methods. A more user-friendly version of TDM architecture will be published for the betterment of space propulsion research. Therefore, suggestions and further recommendations are expected from the domain experts and researchers of this field.

Acknowledgments

This work is assisted and utilized by the space systems experiment facility of the Jet Propulsion Laboratory situated in School of Aerospace Engineering at Beijing Institute of Technology from 2016 to 2020 during the doctoral and masters study tenure of the authors. All respective credit goes to the lambastes, professors, and reviewers to conduct the experiments and other related research and development processes. This work is also grateful to them who provided guidance and impetus to conduct this multidisciplinary research.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data are available on request from the corresponding author upon reasonable request.

Author Contribution Statement

Noman Hossain: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Guorui Sun:** Validation, Investigation, Writing – original draft, Writing – review & editing. **Farjana Islam:** Software, Writing – original draft, Writing – review & editing.

References

[1] Burton, R. L., & Turchi, P. J. (1998). Pulsed plasma thruster. *Journal of Propulsion and Power*, *14*(5), 716–735. https://doi.org/10.2514/2.5334

[2] Li, X. (2022). Gaseous switching arc and extension to pulsed plasma. In *IEEE 67th Holm Conference on Electrical Contacts,* 1–12. https://doi.org/10.1109/HLM54538.2022.9969803

[3] Sun, G., Wu, Z., Li, H., & Zeng, L. (2019). Discharge voltage characteristic in ablative pulsed plasma thrusters. *Aerospace Science and Technology*, *86*, 153–159. https://doi.org/10.1016/j.ast.2019.01.017

[4] Cho, M. G., & Sung, H. G. (2019). Numerical modeling of pulsed plasma thruster performance with Teflon ablation and ionization. *International Journal of Aeronautical and Space Sciences*, *20*(2), 415–422. https://doi.org/10.1007/s42405-018-00139-0

[5] D'Andrea, D., Munz, C. D., & Schneider, R. (2010). Modeling of long-range intra- and inter-species charged particle collisions for PIC simulations. *Communications in Computational Physics*, *7*(5), 877–903. https://doi.org/10.4208/cicp.2009.09.094

[6] Yang, L., Liu, X., Guo, W., Wu, Z., & Wang, N. (2010). Simulation of parallel-plate pulsed plasma Teflon® thruster

based on the electromechanical model. In *2nd International Conference on Advanced Computer Control*, 148–151. https://doi.org/10.1109/ICACC.2010.5487276

[7] Markusic, T., Berkery, J., & Cooley, J. (2001). *Physics and dynamics of current sheets in pulsed plasma thrusters*. US Air Force Office of Scientific Research. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=43d11beebcde1375168094d10e128ff7f97deb30

[8] Yin, L., & Zhou, J. (2007). The performance analyse of the pulsed plasma thruster. In *New Trends in Fluid Mechanics Research: Proceedings of the Fifth International Conference on Fluid Mechanics*, 702–705. https://doi.org/10.1007/978-3-540-75995-9_234

[9] Mashidori, H., Kakami, A., Muranaka, T., & Tachibana, T. (2006). A coaxial pulsed plasma thruster using chemical propellants. *Vacuum*, *80*(11–12), 1229–1233. https://doi.org/10.1016/j.vacuum.2006.01.041

[10] Schönherr, T., Nees, F., Arakawa, Y., Komurasaki, K., & Herdrich, G. (2013). Characteristics of plasma properties in an ablative pulsed plasma thruster. *Physics of Plasmas*, *20*(3), 033503. https://doi.org/10.1063/1.4794198

[11] Kumar, S., Tiwari, P., & Zymbler, M. (2019). Internet of Things is a revolutionary approach for future technology enhancement: A review. *Journal of Big Data*, *6*(1), 111. https://doi.org/10.1186/s40537-019-0268-2

[12] Ciaralli, S., Coletti, M., & Gabriel, S. B. (2016). Results of the qualification test campaign of a Pulsed Plasma Thruster for Cubesat Propulsion (PPTCUP). *Acta Astronautica*, *121*, 314–322. https://doi.org/10.1016/j.actaastro.2015.08.016

[13] Javaid, M., & Haleem Khan, I. (2021). Internet of Things (IoT) enabled healthcare helps to take the challenges of COVID-19 pandemic. *Journal of Oral Biology and Craniofacial Research*, *11*(2), 209–214. https://doi.org/10.1016/j.jobcr.2021.01.015

[14] Rodrigues, D., Carvalho, P., Rito Lima, S., Lima, E., & Lopes, N. V. (2022). An IoT platform for production monitoring in the aerospace manufacturing industry. *Journal of Cleaner Production*, *368*, 133264. https://doi.org/10.1016/j.jclepro.2022.133264

[15] Xu, J., Gu, B., & Tian, G. (2022). Review of agricultural IoT technology. *Artificial Intelligence in Agriculture*, *6*, 10–22. https://doi.org/10.1016/j.aiia.2022.01.001

[16] Khorasani, K., & Valdes, A. (2008). Dynamic neural network-based Pulsed Plasma Thruster (PPT) fault detection and isolation for the attitude control system of a satellite. In *IEEE International Joint Conference on Neural Networks,* 2689–2695. https://doi.org/10.1109/IJCNN.2008.4634175

[17] Tenney, A. S., Glauser, M. N., Lewalle, J., & Ruscher, C. J. (2018). A deep learning approach to jet noise prediction. In *AIAA Aerospace Sciences Meeting*. Advance online publication. https://doi.org/10.2514/6.2018-1736

[18] Hossain, N., Wang, N., Sun, G., Li, H., & Wu, Z. (2020). A reliable data-driven model for Ablative Pulsed Plasma Thruster. *Aerospace Science and Technology*, *105*, 105953. https://doi.org/10.1016/j.ast.2020.105953

[19] Pan, J., Liu, Y., Zhang, S., Hu, X., Liu, Y., & Shao, T. (2023). Deep learning-assisted pulsed discharge plasma catalysis modeling. *Energy Conversion and Management*, *277*, 116620. https://doi.org/10.1016/j.enconman.2022.116620

[20] Plyashkov, Y. V., Shagayda, A. A., Kravchenko, D. A., Lovtsov, A. S., & Ratnikov, F. D. (2022). On scaling of Hall-effect thrusters using neural nets. *Journal of Propulsion and Power*, *38*(6), 935–944. https://doi.org/10.2514/1.B38592

[21] Sun, G., & Wu, Z. (2018). Ignition mechanism in ablative pulsed plasma thrusters with coaxial semiconductor spark plugs. *Acta Astronautica*, *151*, 120–124. https://doi.org/10.1016/j.actaastro.2018.06.027

How to Cite: Hossain, N., Sun, G., & Islam, F. (2025). Data Science and Machine Learning Processes for IoT-Based Pulsed Plasma Thruster Research. *Journal of Data Science and Intelligent Systems*, *3*(3), 231–239. https://doi.org/10.47852/bonviewJDSIS32021517