Enhancing STEM Education through Data Science: Predicting Electric Motor Performance with Linear and OLS Regression

Sundeep H Deulkar, Sagarika Pramod

Pillai College of Engineering, Dr. K. M. Vasudevan Pillai Campus, Sector 16, New Panvel-410206

Corresponding author: Sundeep H Deulkar, Email: saisundeep@mes.ac.in

This study bridges practical engineering tasks with data science analysis to enhance STEM education for high school students. In a hands-on exercise involving 60 teams of 12th-grade students, participants constructed electric motors to achieve maximum rotations per minute (RPM) using various values of magnetic field strength (B), current (I), number of coil turns (N), and coil cross-sectional area (A). We utilized this data to build and train linear regression and Ordinary Least Squares (OLS) regression models to predict RPM and power based on the measured values of B, I, N, and A. The analysis involved cleaning the data, performing regression, and evaluating the model's performance using metrics such as R^2 values, p-values, Mean Squared Error (MSE), and model comparison criteria including AIC and BIC. The linear regression model for RPM yielded an R^2 value of 0.209 and a Mean Squared Error of 1043.14, indicating a weak fit. The OLS regression model for RPM also showed a low R^2 value of 0.205, with an adjusted R^2 of 0.046, and a non-significant F-statistic (p-value = 0.309). In contrast, the linear regression model for power resulted in an R^2 value of 0.450 and an MSE of 1090.11. The OLS regression model for power demonstrated a stronger fit with an R^2 value of 0.573, an adjusted R^2 of 0.487, and a significant F-statistic (p-value = 0.00136), along with lower AIC and BIC values compared to the RPM model. This approach not only facilitated practical learning but also demonstrated the application of data science in analysing and optimizing engineering designs. The results highlight the critical factors influencing electric motor performance and underscore the educational value of integrating data- driven analysis into STEM projects. This study provides valuable insights for future curriculum development, emphasizing the role of data science in enhancing experiential learning.

Keywords: Electric Motor, Experiential Learning, Linear Regression, Predic- tive Modeling.

2025. In Sandeep Kumar & Kavita Sharma (eds.), *Computational Intelligence and Machine Learning*, 111–120. Computing & Intelligent Systems, SCRS, India. DOI: https://doi.org/10.56155/978-81-975670-5-6-9

1 Introduction

1.1 A Context and Background: Integrating Data Science with STEM Education

The integration of data science into STEM (Science, Technology, Engineering, Mathematics) education represents a critical evolution in teaching methodologies that aims to prepare students for the increasingly data-driven world. Traditionally, STEM education has focused on theoretical knowledge and isolated practical applications within each discipline. However, with the advent of big data and advanced analytics, there is a growing need to equip teachers with skills that cut across these disciplines and prepare students for real-world challenges and application based projects.

Data science, with its emphasis on data collection, analysis, and interpretation, enhances the teaching-learning experience by providing a practical, hands-on approach to solving complex problems. Integrating data science into STEM education encourages teachers to engage students in interdisciplinary learning, where they (teachers and later students) can apply mathematical and computational methods to scientific and engineering problems. This integration not only deepens the teacher's understanding of core STEM concepts but also cultivates critical thinking, problem-solving, and analytical skills.

Educational frameworks and studies, such as those by Blikstein (2018)[1] and Li et al. (2020)[2], underscore the transformative potential of integrating data science with STEM education. They highlight how data-driven approaches can foster a more engaging and effective learning environment, helping students to see the relevance of their studies in real-world contexts.

1.2 Objective of the Study:

The primary objective of this study is to explore the pedagogical benefits of integrating data science with hands-on STEM education. Specifically, the study aims to:

Design and Execute an Educational Exercise: Engage students in constructing electric motors using various configurations of magnetic field strength (B), current (I), number of coils turns (N), and coil cross-sectional area (A) to achieve optimum rotations per minute (RPM).

Collect and Analyse Experimental Data: Utilize the data collected from these experiments to build and train a linear regression model.

Predict Motor Performance: Use the model to predict RPM and power output, thereby demonstrating the practical application of data science techniques in analysing and optimizing engineering designs.

In conclusion, integrating data science with STEM education not only enriches the learning experience but also equips students with the skills needed to thrive in a data- centric world. This study aims to demonstrate the practical benefits of such an inte- grated approach and provide a framework for future educational innovations.

2 Experimental Setup

2.1 Experimental Task given to the Students

The Problem Definition: To design an operating model of Electric Motor with optimum use of materials provided for achieving Maximum Rotations Per Minute (RPM). Specific Instructions: Once students have decided the no of turns, they cannot increase it but they had scope of decreasing it. Two sets of coils with different gauges 26 and 29 were provided. Student could choose any one of the gauges.

Student have to measure the area (A) of the wound coil with the help of the ruler provided.

Students were provided with 2 types of axles/shafts (a round pencil and a chop stick) and they had to choose one.

Students were also provided with two types of pairs of magnets (small and medium). The alignment of magnets was their choice. In addition, a 12 V power supply was pro- vided.

With the optimum use of these components, students had to achieve the highest RPM (Rotations per minute) of the motor.

Judge's sheet: The Judge's sheet was designed to collect the following data:

1. Number of turns of coil
2. Area of the coil
3. Type of axle used
4. Total weight of Axle and coil.
5. Type of magnet (Medium/small)
6. Distance between magnet pole pieces and coil center
7. RPM value as measured by laser tachometer.

Figure below shows a snap-shot of a working model of the electric motor, made by one of the groups of students.

Figure1. Actual working model of the electric model

2.2 Data Acquisition

Magnetic field B: The value of B was calculated by first calculating the pole strength of the magnet pairs (small as well as medium) by measuring the magnetic fields in between the magnet pairs (separated by a fixed distance) using a gauss meter. For individual pairs the distance varied and hence the magnetic field was deciphered by knowing the magnets pole strength and the distance between the axle and the magnet pole, using the following formula:

$$B = \frac{\mu_0 m_s}{2\pi r^3} \qquad (1)$$

Where μ_0 is the magnetic permeability of free space. And ms is the magnetic pole strength r^3 is the cube of the distance of the axle from the magnetic pole.

Current I: This is calculated using Ohm's law:

$$I = \frac{V}{R} \tag{2}$$

Where V is the voltage (12 V) and R is the resistance of the coil wire. This value is calculated using the following formula:

$$R = \frac{\rho l}{A} \tag{3}$$

Where ρ is the resistivity of copper ($1.7 \times 10^{-8}\Omega$m), l=length of the coil wire, and A is the cross-sectional area of the wire

$$A' = \pi r^2 \tag{4}$$

Where r is the radius of cross section of the wire.

Total number of turns N: This value was obtained from the Judge's sheet.

Cross-sectional Area of the coil A: This value was obtained from the Judge's sheet and was also measured separately.

Figure 2. below displays the histograms for the magnetic field strength (B), current (I), number of turns (N), and coil area (A). Each histogram illustrates the distribution of values for the respective variable, providing insights into their frequency and range within the dataset.

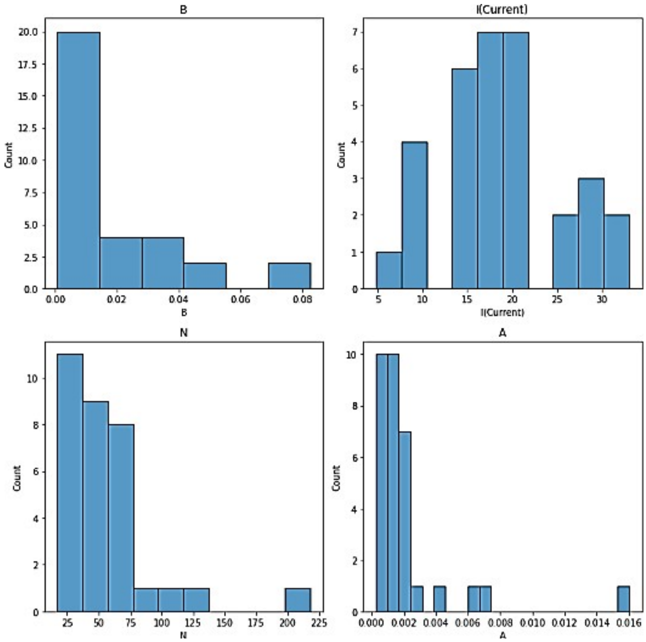

Figure 2. Distribution Histograms of Key Variables: B, I, N and A.

RPM values: These values were obtained by gluing a reflecting paper on a plastic propeller attached to the motor axle and observing the value of the Rotations per Minute displayed by a standard laser tachometer.

2.3 Data Processing

The target variable of Torque 'T' was calculated using the following formula [3]:

$$T = B \times I \times N \times A \tag{5}$$

Where B, I, N and A are the independent variables.
The value of ω, the angular velocity was calculated from the RPM values using the following formula:

$$\omega = \frac{RPM * 2 * \pi}{60} \tag{6}$$

Finally the second target variable —power P, was estimated using the formula [3]:

$$P = T \times \omega \tag{7}$$

Instances where the motor did not function were discarded and instances where the motor showed a slight rotation were assigned an arbitrary RPM value of 0.5. So although 60 teams participated only 25 teams had a working motor.

3 Principles of Linear Regression

The multiple regression model [4-6] aims to establish a linear relationship between the dependent variable Y (RPM/ ω or power) and the independent variables Xi (B, I, N, A). The general form of the multiple regression equation is:

$$Y = \beta_0 + \beta_1 B + \beta_2 I + \beta_3 N + \beta_4 A + \varepsilon \tag{8}$$

Where :

- Y is the dependent variable (RPM or power).
- β_0 is the intercept of the regression line.
- β_1, β_2, β_3, β_4 are the coefficients corresponding to the independent variables B, I, N, and A, respectively.
- ϵ is the error term, which accounts for the variability in Y that cannot be explained by the linear relationship with the independent variables.

3.1 Fitting the Model Data Collection:

Measure the values of the independent variables B, I, N, and A for each motor constructed by the students.
Record the corresponding RPM (and hence ω) and power (P) outputs.

Building the Model:
Use a programming language like Python (with libraries such as scikitlearn or stats- models) to fit the multiple regression model.
Input the collected data to estimate the coefficients β_0, β_1, β_2, β_3, β_4 using the method of least squares, which minimizes the sum of the squared differences between the ob- served and predicted values of the dependent variable.

Interpretation of Coefficients:
Intercept ($\beta 0$): Represents the predicted value of RPM or power when all independent variables (B, I, N, A) are zero. While this might not have a practical physical interpretation, it serves as a baseline in the model.
Coefficient of B ($\beta 1$): Indicates the change in the dependent variable (RPM or power) for a one-unit increase in the magnetic field strength (B), holding all other variables constant.
Coefficient of I ($\beta 2$): Represents the change in the dependent variable for a one-unit increase in the current (I), holding all other variables constant.
Coefficient of N ($\beta 3$): Reflects the change in the dependent variable for a one-unit increase in the number of coil turns (N), holding all other variables constant.
Coefficient of A ($\beta 4$): Shows the change in the dependent variable for a one-unit in- crease in the cross-sectional area of the coil (A), holding all other variables constant.

Application to Electric Motor Data Case:
After fitting the model, one can use the estimated coefficients to predict the RPM and power for new configurations of B, I, N, and A. This involves plugging in the values of these variables into the regression equation:

$$PredictedRPM = \beta_0 + \beta_1 B + \beta_2 I + \beta_3 N + \beta_4 A + \varepsilon \tag{9}$$

$$Predicted\ Power = \beta_0' + \beta_1' B + \beta_2' I + \beta_3' N + \beta_4' A + \varepsilon \tag{10}$$

While the form of the equations is the same, the regression coefficients (βi) are different for each dependent variable. Each set of coefficients is calculated to best fit the data for RPM and power, respectively.

By doing so, one can evaluate how changes in the design parameters (B, I, N, A) affect the performance of the motors. This analysis helps in understanding the relationship between the variables and optimizing the motor designs for better performance.

Steps in Python code:

1 **Loading Data**: The data is loaded from a CSV file using pd.read_csv(). Make sure to replace 'path_to_file.csv' with the actual path to the CSV file [7].
2 **Independent Variables**: Columns B, I (Current), N, and A are used as independent variables.
3 **Dependent Variables**: Columns RPM and P are used as dependent variables for the two models. **Splitting Data**: The data is split into training and testing sets for both RPM and power predictions.
4 **Training Models**: Two linear regression models are trained separately for RPM and power using sklearn’s Linear Regression. Two separate linear regression models are trained—one for predicting RPM and another for predicting power. Each model will has its own set of coefficients.
5 **Displaying Coefficients**: The coefficients for each model are displayed us- ing pandas Data Frames.
6 **Detailed Model Summary**: stats models is used to fit the models again and display detailed summaries. The values of R^2 and p-values for each model are calculated to evaluate model performance and the significance of the predictors.

R^2 (Coefficient of Determination): Measures the proportion of the variance in the dependent variable that is predictable from the independent variables. An R^2 value closer to 1 indicates a better fit.

P-Values: Assess the statistical significance of each coefficient. A low p-value (typically < 0.05) suggests that the corresponding independent variable significantly contributes to the model.

4 Result Analysis and Insights:

Based on the latest OLS regression results for RPM and Power, we can derive several insights regarding the performance and predictive power of our models.

4.1 RPM Model Analysis

The table below (Table1.a.) provides a detailed summary of the Ordinary Least Squares (OLS) regression analysis performed to predict the rotations per minute (RPM) of the electric motors based on the magnetic field strength (B), current (I), number of coil turns (N), and coil cross-sectional area (A). This analysis highlights the relationship between the predictors and the response variable, offering insights into the factors influencing motor performance.

Table 1. a.RPM/ωOLS model summary

OLS Regression Results

Dep. Variable:	RPM	R-squared:	0.205
Model:	OLS	Adj. R-squared:	0.046
Method:	Least Squares	F-statistic:	1.286
Date:	Fri, 31 May 2024	Prob (F-statistic):	0.309
Time:	14:33:52	Log-Likelihood:	-202.28
No. Observations:	25	AIC:	414.6
Df Residuals:	20	BIC:	420.7
Df Model:	4		
Covariance Type:	nonrobust		

Linear Regression Results: Mean Squared Error:1043.14 R-squared: 0.209

Key Insights

R-squared Value
The R-squared value of 0.205 (OLS) and 0.209 (linear regression) indicate that approximately 20.5% and 20.9%, respectively, of the variability in RPM can be explained by the model using the independent variables (B, I, N, A). Both suggest a relatively weak fit, as a substantial portion of the variability is not captured by the models.

Adjusted R-squared Value
The adjusted R-squared value of 0.046 (OLS) confirms that the model has limited explanatory power for predicting RPM after adjusting for the number of predictors and sample size.

F-statistic and p-value
The F-statistic of 1.286 and the associated p-value of 0.309 (OLS) indicate that the model is not statistically significant at typical significance levels (e.g., 0.05). This means that the null hypothesis that all coefficients are equal to zero cannot be rejected, suggesting that the predictors do not significantly improve the model fit for RPM.

Mean Squared Error (MSE)
The MSE of 1043.14 (linear regression) indicates the average squared difference between the observed and predicted RPM values. A high MSE further supports the notion of a weak model fit.

AIC and BIC Values
The AIC (414.6) and BIC (420.7) values (OLS) provide metrics for model comparison. Higher values suggest that the model may not be the most efficient in terms of explaining the data while balancing complexity.

Log-Likelihood
The log-likelihood value of -202.28 (OLS) offers a measure of the model fit. Com- paring this with other models can help determine which model better explains the data.

4.2 Power Model Analysis

The table below(Table 1.b.) provides a detailed summary of the Ordinary Least Squares (OLS) regression analysis performed to predict the rotations per minute Power of the electric motors based on the magnetic field strength (B), current (I), number of coil turns (N), and coil cross-sectional area (A). Similar to Table 1.a. this analysis too high- lights the relationship between the predictors and the response variable, offering in- sights into the factors influencing motor performance.

Table 1. b. Power OLS model summary

OLS Regression Results

Dep. Variable:	P	R-squared:	0.573
Model:	OLS	Adj. R-squared:	0.487
Method:	Least Squares	F-statistic:	6.705
Date:	Fri, 31 May 2024	Prob (F-statistic):	0.00136
Time:	14:33:52	Log-Likelihood:	-135.14
No. Observations:	25	AIC:	280.3
Df Residuals:	20	BIC:	286.4
Df Model:	4		
Covariance Type:	nonrobust		

Linear Regression Results
Mean Squared Error:1090.11 R-squared: 0.450

Key Insights

R-squared Value
The R-squared value of 0.573 (OLS) and 0.450 (linear regression) indicate that ap- proximately 57.3% and 45.0%, respectively, of the variability in Power can be explained by the model using the independent variables (B, I, N, A). Both suggest a moderate to strong fit, as a substantial portion of the variability is captured by the models.

Adjusted R-squared Value
The adjusted R-squared value of 0.487 (OLS), while slightly lower, still indicates a reasonable explanatory power for predicting Power. It adjusts for the number of predictors and sample size, affirming the model's strength.

F-statistic and p-value
The F-statistic of 6.705 and the associated p-value of 0.00136 (OLS) indicate that the model is statistically significant. This suggests that the independent variables collectively contribute to the prediction of Power, rejecting the null hypothesis that all coefficients are equal to zero.

Mean Squared Error (MSE)
The MSE of 1090.11 (linear regression) indicates the average squared difference between the observed and predicted Power values. This value helps in assessing the predictive accuracy of the model.

AIC and BIC Values
The AIC (280.3) and BIC (286.4) values (OLS) are lower than those for the RPM model, suggesting a better fit relative to model complexity. These values support the Power model's efficiency in explaining the data.

Log-Likelihood
The log-likelihood value of -135.14 (OLS) indicates a better fit compared to the RPM model. This aligns with the higher R-squared and lower AIC/BIC values, showing that the Power model better captures the underlying data structure.

5 Summary

5.1 RPM Model

Weak Fit: Both OLS and linear regression models explain a small portion of the variability in RPM, indicated by the low R-squared values. This suggests that the factors B, I, N, and A, as applied by the students, have limited predictive power for RPM in the constructed motors. This highlights an opportunity for students to further explore and understand the complexities involved in motor performance.

Statistically Insignificant: The high p-value (OLS) suggests that the predictors are not significantly related to RPM. This result encourages students to delve deeper into the principles of motor physics and the impact of each variable, promoting critical thinking and problem-solving skills.

Model Comparison: Higher AIC and BIC values (OLS) indicate less efficiency com- pared to the Power model. This provides a practical demonstration for students on how model selection criteria can inform the evaluation of model performance and efficiency.

5.2 Power Model

Moderate to Strong Fit: Both OLS and linear regression models explain a substantial portion of the variability in Power, supported by higher R-squared values. This illustrates to students how certain variables have a stronger influence on motor performance, reinforcing their understanding of the underlying physical principles.

Statistically Significant: The low p-value (OLS) confirms that the predictors significantly relate to Power. This outcome emphasizes the importance of proper data collection and analysis in experimental physics and engineering, showing students the practical applications of statistical significance in real-world scenarios.

5.3 Insights

While the high school students participated in the hands-on motor building activity, the subsequent data analysis was performed by the teachers. This division of tasks provided a comprehensive educational experience where students applied their understanding of physics and engineering to construct the motors, while teachers used the collected data to perform detailed statistical analyses.

This approach highlighted the interconnectedness of practical engineering tasks and data science techniques. It underscored the educational value of integrating data-driven analysis into STEM education, demonstrating to the learners how empirical data can be analysed to optimize designs and improve performance, thus promoting a deeper understanding of the scientific method and data science principles in engineering applications. The STEM based approach thus deepens the students' understanding of physics and engineering principles through experiential learning, thus fulfilling the primary objective of this study.

6 Acknowledgements

Dr. Sundeep H Deulkar would like to thank Dr. Arun Pillai, HOD of Applied Sciences, Mathematics and Humanities department of PCE for spearheading the MAKERTHON 2023 event for high school students that involved the task of motor building. This was the source of data for our study. The support of my other colleagues in the engineering physics lab is also deeply acknowledged: Dr. Ravindra Nikam and Dr. Mahendra Khandkar for contributing towards building and testing the prototype mo- tors for the event. Sincere thanks to Lab Incharge Mr. Sudhir A. and Lab assistant Mr. Ankush for their wholehearted support throughout the event and also post event data collection and measurements. I would also like to thank Prof. Binshu Babu and Prof. Amey Nijasure for providing guidance in measuring the magnetic field using a gauss- meter. Finally I would also like to thank Prof. Salim Jafri for assistance in measurement of crossectional areas of wire using digital Vernier calliper and micrometer screw gauge.

References

[1] Blikstein,P: Using Data Science to Transform K-12 STEM Education. Retrieved from the Data Science Institute at Columbia University, 2020/06/01

[2] Li, Y., Schoenfeld, A. H., DiSessa, A. A., Graesser, A., & Benson, L.:Data Science for Enhancing STEM Education. Journal of Educational Psychology, 112(5), 1017-1029 (2020).

[3] Halliday,D,, Resnick,R. Walker,J.: Fundamental of Physics, 7edn. Hoboken, NJ: Wiley(2005)

[4] Montgomery, D. C., Peck, E. A., & Vining, G. G.: Introduction to Linear Regression Analysis. 6th edn. Wiley (2012).

[5] Freedman, D. A.: Statistical Models: Theory and Practice. Cambridge University Press (2009).

[6] Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. : Applied Linear Statistical Models. 5th edn. McGraw-Hill (2005).

[7] Rao,N: Core Python Programming. 2nd edn. Dreamtech,(2018).