The Investigation of Logistic Regression Methods Applied to Engineering Education using Project Based Learning for Airport Systems Design

Willie L. Brown, Jr.
Department of Engineering and Aviation Sciences
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
wlbrown@umes.edu

Ibibia Dabipi
Department of Engineering and Aviation Sciences
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
ikdabipi@umes.edu

Dinesh Sharma
Department of Business, Management and Accounting
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
dksharma@umes.edu

Lei Zhang
Department of Engineering and Aviation Sciences
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
lzhang@umes.edu

Weiwei Zhu-Stone
Department of Computer Science and Engineering Technology
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
wzhu@umes.edu

Lanju Mei
Department of Engineering and Aviation Sciences
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
lmei@umes.edu

Urban T. Wiggins
Division of Data Science and Visualization
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
utwiggins@umes.edu

Tiara T. Cornelius
Department of Mathematics
University of Maryland Eastern Shore
Princess Anne, Maryland, USA
tdturner@umes.edu

Sherley Jones, Jr.
Wallops Flight Facility, Goddard Space Flight Center
National Aeronautics and Space Adminiistration
Wallops Island, Virginia, USA
sherley.l.jones@nasa.gov

Roland Wescott
Wallops Flight Facility, Goddard Space Flight Center
National Aeronautics and Space Adminiistration
Wallops Island, Virginia, USA
roland.wescott-1@nasa.gov

J. Anthony Sharp
Director of Aviation Science (Retired)
Elizabeth City State University
Elizabeth City, North Carolina, USA
jasharp1234@gmail.com

Furman Glenn
E.I. DuPont de Nemours & Company
Retired Senior Scientist at DuPont Performance Elastomers
Louisville, Kentucky, USA
glenccu@gmail.com

***Abstract*—This is a work in progress on the research-to-practice to advance engineering development using logistic regression design methods in engineering education through project-based learning (PBL) activities. The aviation community has witnessed a growing concern with safety and security issues due to system vulnerability of manned and unmanned aircrafts. The implementation PBL in engineering education with major airports identified the benefits of this investigation as a multi-disciplinary approach to address system vulnerabilities and analysis design approaches. This investigation includes defined methods and the factors to assess undergoing efforts of environment threats associated with complex networked systems in aviation through analysis design activities with machine learning. The ability to examine the analysis design allows learners in engineering education to deploy methods that will assess the combination of techniques and security consideration through machine learning. This project comprised of a collected dataset of network traces and malicious traffic that requires a solution and strategic model based on various scenarios to improve the software architecture design. Aviation safety and security standards are critical to the degree in which threats are constituted as each node within the network determined specific case studies for evaluation. This approach will feature a dataset that assess the spam detection categorization from a human behavior perspective and the factors to formulate data mining techniques and models in R programming language. The framework adopts the use of logistic regression and proposed techniques of the performance baseline to explore a systematic approach in the engineering design process. PBL in engineering education revealed a viewpoint from interdisciplinary approach as this study will have a wide range of features in the design analysis (e.g., revealing unwanted electronic information spread causing concerns and environmental threats to the aviation community). This study is developed to examine a common framework for processing and structuring a model using logistic regression analysis to classify key elements in the engineering process and PBL activities. (*Abstract*)**

Keywords—Project Based Learning, Engineering Education, Airport Safety, Logistic Regression Analysis, Research to Practices, Design Methods, R Programming Language *(key words)*

I. Introduction

The ability to examine education effectiveness and integration of project-based learning (PBL) allows for the advancement of engineering education to further explore specific practices using specific methods. This approach introduces learning outcomes in key disciplines by building on problems identified in research practices. In this study, logistic regression was integrated to address the various parameters to perform course-learning activities according to the research practices in airport environments for engineering education. The practice in education highlights research modeling and methods to examine challenges and explore relationships with regard to STEM education and industry developments. The specific engineering courses included both lower and upper levels education learning to advance research practices. The theory with the use of logistic regression methods, the interaction with industry officials to examine the course learning objectives using PBL activities. The community interaction with the airport environments aided the education measures to be defined based on the overall learning outcomes in engineering courses. The analysis and results will include the investigation of STEM practices to explore course-learning objectives and provides a lifelong learning component with delivery methods in engineering education. From the results, the study examines engineering course learning outcomes to integrate PBL requirements into the airport practices. Our findings supported these core parameters to bring awareness and a PBL connection with STEM courses for the advancement using logistic regression methods.

II. Scope

The assessment of safety methods with airport practices such as spam detection has been a significant factor to address undergoing efforts in order to examine environmental threats associated with complex networked systems [1]. The ability to deploy methods that will assess the combination of techniques of network for security the various consideration will build on the degree in which threats are constitute as a number of nodes within a complexed network [1]. These conditions comprised of a collected dataset of network traces and malicious traffic that requires solutions and strategies to model attack scenarios and specific case studies. The study will be essential for both cyber security and cyber defense application using statistical data mining techniques for airport practices. This approach will feature a dataset concerning spam detection to evaluate the model using logistic regression [2]. The rationale of this type of analysis is to assess the methods and spam detection of relationship between variables from a perspective spam e-mails based on statistical data mining techniques and models in R language.

III. Logistic Regression Applied to Supervised Learning For Data Collection: Background And Literature Review

Logistic regression methods and approaches are applied to better understand PBL performance and evaluate datasets using specific methods. These methods are usually applied to a PBL application using datasets (e.g., sample condition measures, accuracy, and structures) as a tool. The modeling of specific techniques and approaches are effective tools to evaluate and model the proposed method using supervised learning environments. The study adopted supervised learning as an approach to evaluate application environments and conditions according to the classification [2]. The implementation requirements to train the model supports this empirical approach of the results from different training processes and techniques conducted in supervised learning. However, the level of predication and related experiments also corresponds with the ability to conduct supervised training and the algorithm mapping of function (e.g., input and output relationships). Whereas the unsupervised learning presents, a structure in the data without any corresponding outputs [2][3].

The approach to implement such model for supervised learning using the techniques mentioned requires the proposed conditions and strategies to be designed and introduced as a simple criterion. An investigation of PBL performance indicates that the leverage of a simple metric can be utilized to explore patterns, which helps to identify the probability of distribution in class labels (i.e., allows for data characteristics of predictors) [3][4]. From the PBL findings, the studies were able to perform related learning methods and approaches to understand the level of accuracy in performance. This metric generated an understanding to apply with a level of accuracy in measures and classification for further investigation among relationships [2][4]. As such, the approach will yield to a logistic regression and the training data will be available to build a prediction [1][3]. This implementation of specific approaches in PBL provides uniqueness regarding datasets and behavioral patterns that may not exist due to the emergence of parameters and schemes. Therefore, the information to analyze how to model PBL engineering education performance will be inherently characterized and this study illustrates the unique behaviors of investigating broad statistical tools using logistic regression approaches [2][5].

IV. Logistic Regression Methods and The Tool For Assessment: PBL in Engineering Education and The Data Preprocessing Practices Explored By Students in the Design Process

The data analysis explains a quantitative assessment regarding the theory properties of logistic regression and the prediction using probability with represent to explanatory variables. As such, the analysis will support the view that the user observation of spam email detection process and the techniques to address knowledge discoveries for data preprocessing of engineering students. The purpose of this analysis is to utilize the design approach of the content specific knowledge from the dataset to examine the data preprocessing concepts involving the collection of spam emails. In creating a systematic approach using logistic regression method in engineering education for assessment, the practices identified strategies to improve reliability related to outcomes and education objectives. The following three (3) indicators explored the relationships among data preprocessing actives and implementation strategies, which include: a) completeness; b) time-related (and data variation); and c) correctness [3]. These relationships amongst data preprocessing activities investigated the data cleaning practices (e.g., completeness, data variation, and correctness) of the engineering students to assess data handling. The data handling to understand validity supports the students' ability to treat missing values and detecting outliers in the areas that involved [3]: a) assessment of datasets and the volume for redundancy to ensuring effectiveness; b) determine whether the volume and data values meets the performed threshold for assessment to include in the engineer design process; and c) deployment of data cleaning techniques to support R programming language (i.e., implementation of a tool to evaluate data variation and statistical computing for reliability).

R programming language was used by the engineering students to develop a framework for processing any statistical operations according to the engineering design. The goal and objective with using R language as the statistical tool allowed to generate techniques and performances appropriate for theory of logistic regression with regard to the data preprocessing practices proposed. This assessment enabled the decision making process to explore and facilitate a systematic approach to detect spam and potential misleading features of product reviews. Spam content is generated and the information repository created concerns due to the unwanted electronic information spread in the airport environments [4].

V. Theorectical Framework And Methodology: PBL Research To Practice With Engineering Education

Ciuhu et al., (2017) created a preprocessing approach based on the logistic regression techniques that allowed for practical usability to understand the parameters and the influence with regard to the describe data and relationships. The theoretical framework of this study had considered steps to define the specific characteristics to model the relationships and predictors to compare the parametric models as part of the class of generalized linear model (GLM) [5]. This methodology to develop this approach includes four (4) criteria and activities:

- Identify the critical for proper interpretation of the results to consider the justification and algorithms used to assessed performance;
- Evaluate the model to understand the trained using quantifiers and relationships, in particular to issues involving probabilistic forecasts;
- Examine the thresholds and the model predictors pertaining to the correlated variables, which can result in overfitting to address the issues required in preprocessing; and
- Determine the preprocessing steps in order to understand the model predictors (e.g., undergo dimensionality reduction) that could result to exploring principal components analysis (PCA) [5].

These steps will be integrated to complete the preprocess phase of the study in order to obtain and determine each condition and task to evaluate the models and trained to forecast. The implementation design will assess the role of the preprocessing approach and the performance activities for integration (see figure one below of the summary for the descriptive statistics).

```
Call:  glm(formula = spam ~ ., family = binomial, data = train.data)

Coefficients:
             (Intercept)            word_freq_make         word_freq_address                word_freq_all
              -1.524e+00                -5.228e-01                -1.540e-01                    4.653e-02
            word_freq_3d             word_freq_our            word_freq_over             word_freq_remove
               2.317e+00                 5.978e-01                 1.360e+00                    2.136e+00
      word_freq_internet           word_freq_order            word_freq_mail            word_freq_receive
               7.278e-01                 5.761e-01                 2.265e-01                   -2.122e-02
          word_freq_will          word_freq_people          word_freq_report          word_freq_addresses
              -1.738e-01                -1.172e-01                 1.882e-01                    1.085e+00
          word_freq_free        word_freq_business           word_freq_email                word_freq_you
               1.251e+00                 1.036e+00                 2.198e-01                    6.082e-02
        word_freq_credit            word_freq_your            word_freq_font                word_freq_000
               8.999e-01                 1.886e-01                 2.146e-01                    2.694e+00
         word_freq_money              word_freq_hp             word_freq_hpl             word_freq_george
               3.602e-01                -2.377e+00                -1.042e+00                   -8.484e+00
           word_freq_650             word_freq_lab            word_freq_labs             word_freq_telnet
               1.057e+00                -2.535e+00                -3.686e-01                   -1.328e-01
           word_freq_857            word_freq_data             word_freq_415                 word_freq_85
               4.209e+00                -6.488e-01                 6.490e-01                   -1.896e+00
    word_freq_technology            word_freq_1999           word_freq_parts                 word_freq_pm
               9.700e-01                 6.744e-02                -6.941e-01                   -4.448e-01
        word_freq_direct              word_freq_cs         word_freq_meeting           word_freq_original
              -4.791e-01                -4.417e+01                -2.189e+00                   -8.359e-01
       word_freq_project              word_freq_re             word_freq_edu              word_freq_table
              -1.536e+00                -8.986e-01                -1.301e+00                   -1.617e+00
    word_freq_conference               char_freq_.            char_freq_..1                char_freq_..2
              -3.316e+00                -1.259e+00                -1.460e-01                   -3.948e-01
           char_freq_..3             char_freq_..4             char_freq_..5   capital_run_length_average
               4.248e-01                 5.417e+00                 1.945e+00                    4.985e-03
capital_run_length_longest   capital_run_length_total
               6.822e-03                 8.409e-04

Degrees of Freedom: 3240 Total (i.e. Null);  3183 Residual
Null Deviance:      4363
Residual Deviance: 1272        AIC: 1388
```

Fig. 1. Model Output of the Coefficients and Residual Deviance in R Language to Examine the Theoricatial Framework and Methodology

VI. Results And Findings: The Implementation of Research Performance Models With Logistic Regression Methods in Engineering Education

The performed practices to build on the preprocessing activities will adopt the criteria identified related to the approach and steps presented. Therefore, the data set characteristics call as a GLM represented 3240 total degrees of freedom and 1272 residual deviance [6]. According to the summary model (output of the coefficients, p value, and standard error for each independent variable and intercept), the null deviance was 42362.9 on 3240 degrees of freedom and residual deviance was 1272.0 on 3183 degrees of freedom with a number of fisher scoring iterations value of 13. Similar examples found to complete step (3) were highlighted using the benefits of comparative analysis performance to understand measures described on parameters in defining critical points to demonstrate cross validation and the accuracy of classification [5]. The defined critical points of concern to scale the data and handle missing values demonstrated as a baseline classifier involved evaluation measures (see table one below).

Data Preprocessing Results	**Engineering Students: Implementation Strategies and Approach to Apply the Design Performance Models using Logistic Regression Methods**
Data transformation actions to address data type conversions, normalization, and scaling using the logistic regression methods.	Step 1: Engineering students aligned and evaluated the dataset relationships with design requirements based on distribution to handle parameters for data transformation [5]
	Step 2: Engineering students directly mapped and documented the information to re-characterize the implementation strategies with previously data transformation results [5]
	Step 3: Engineering students analyzed and decided on the performance requirements to build on the efficiency for suitability as such target datasets to support any assumptions [5]

Table 1. Engineering Students' Steps to Implement the Logistic Regression Analysis of the System Design for Performance Modeling

The logistic regression methods and results were best served to determine the coefficients and an intercept to handle parameters and conditions for the comparative analysis performance. Figure two below describes the results using the model output of the coefficients and an intercept for data preprocessing in R (i.e., target versus non-target techniques in the performance model to introduce in the engineer design).

```
> exp(coef(model))
             (Intercept)            word_freq_make         word_freq_address                word_freq_all
            2.178965e-01              5.928730e-01              8.573029e-01                 1.047632e+00
            word_freq_3d             word_freq_our            word_freq_over             word_freq_remove
            1.014905e+01              1.818134e+00              3.894804e+00                 8.466111e+00
      word_freq_internet           word_freq_order            word_freq_mail            word_freq_receive
            2.070542e+00              1.779153e+00              1.254249e+00                 9.789995e-01
          word_freq_will          word_freq_people          word_freq_report          word_freq_addresses
            8.404767e-01              8.894277e-01              1.207112e+00                 2.959709e+00
          word_freq_free        word_freq_business           word_freq_email                word_freq_you
            3.494542e+00              2.816825e+00              1.245889e+00                 1.062703e+00
        word_freq_credit            word_freq_your            word_freq_font                word_freq_000
            2.459254e+00              1.207534e+00              1.239380e+00                 1.478994e+01
         word_freq_money              word_freq_hp             word_freq_hpl             word_freq_george
            1.433628e+00              9.281782e-02              3.528144e-01                 2.067503e-04
           word_freq_650             word_freq_lab            word_freq_labs             word_freq_telnet
            2.878298e+00              7.929300e-02              6.917217e-01                 8.756260e-01
           word_freq_857            word_freq_data             word_freq_415                 word_freq_85
            6.725866e+01              5.226483e-01              1.913642e+00                 1.502018e-01
    word_freq_technology            word_freq_1999           word_freq_parts                 word_freq_pm
            2.638040e+00              1.069763e+00              4.995426e-01                 6.409329e-01
        word_freq_direct              word_freq_cs         word_freq_meeting           word_freq_original
            6.193628e-01              6.534420e-20              1.119875e-01                 4.334770e-01
       word_freq_project              word_freq_re             word_freq_edu              word_freq_table
            2.152751e-01              4.071557e-01              2.722501e-01                 1.985108e-01
    word_freq_conference               char_freq_.             char_freq_..1                char_freq_..2
            3.629136e-02              2.840061e-01              8.641685e-01                 6.738422e-01
           char_freq_..3             char_freq_..4             char_freq_..5   capital_run_length_average
            1.529360e+00              2.252099e+02              6.991903e+00                 1.004998e+00
capital_run_length_longest  capital_run_length_total
            1.006845e+00              1.000841e+00
```

Fig. 2. Results using the Model output of the Coefficients and an Intercept for data preprocessing in R

A. *Findings: PBL Benefits in Engineering Education using Logistic Regression Methods with R Programming Language*

The findings in this study were explained and interpreted by the engineering students based on the model outputs and areas presented in the data collection. In their study, the following was examined from the first (10) estimated values as a result (this approach highlighted the confusion matrix for the training set as executed - see figure three). In addition to the confusion matrix, the estimated values were examined and assessed using the variable predictions for the test data as shown in figure four (4). These outcome measures are determined from predictions of the confusion matrix for the test data findings and to explore further implications. The findings from the engineering students discovered the plotted results and the development of residuals and predictions to assess the implications of PBL in the design process (see figures below).

Confusion Matrix (Training Set Data)	0	1
0	1857	135
1	87	1162

Fig. 3. Results of the Confusion Matrix for the Training Set Evaluated in R Programming Langauge

Confusion Matrix (Test Data)	0	1
0	799	49
1	45	467

Fig. 4. Findings of the Confusion Matrix for the Test Data with regard to the Prediction Response in R Programming Langauge

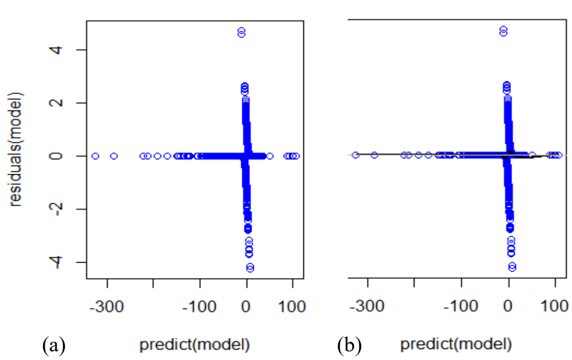

Fig. 5. Plot Comparision: (a) Findings of the Results and the Association of the Models with regard to Residual and Predict Values; (b) Detail Analysis Model: Performance of the Lowess Function to Plot Findings of the Results and the Association of Residual and Predict

The data findings allowed for the evaluation of the models and the test sets associated with aviation safety practices to use within the engineering design based on specific factors. The factors of "0" were identified as *No Spam* in comparison to a factor of "1" resulted to *Yes, it is Spam*. The performance model of predications provided training sets and test sets with a evaluation metrics as presented in figures 3 through 5. These findings are vital to understand and determine the performance measures with correlated predictors using simplified parameters. The investigation of this technique from the engineering students' perspective provided meaningful benefits to model and explore pedagogical practices that support: a) PBL findings to examine the relationships perceived to system design and features according to the engineering requirements; b) PBL application conditions and the design process to test performance measures for system connections; and c) PBL education objectives to align the findings and evaluation process to measure design conditions by using regression analysis for predicting outcomes (or explaining relationships not causation) based on the system requirements.

VII. Conclusion and Future Implications

In conclusion, project-based learning in the classroom environment explored the heart of engineering practices and the research for airport security and safety environments. The proposed concept in PBL was designed to demonstrate the theory of logistic regression methods with engineering practices for research development. The PBL design included core applications to support several underlying features and explored new discoveries that will be beneficial engineering education. The integration of R language to perform an assessment model for logistic regression methods had increased the understanding of research-to-practices, which created new discoveries. The diverse range of airport services and the increasing number of threats for engineers' to demand defined the airport environment and safety challenges. This approach presented a prototype that includes a methodological approach in aiding airports with information and services rendered for PBL improvement. The PBL factors and efficiency were the primary drivers in this study. In order to achieve this concept, the selected study explored the user generated content for analysis to identify and assess the classification process presented with the logistic regression theory. The learning properties in modeling with comparison of results and output measures examined the methods of monitoring, identifying, and filtering out unwanted practices [7]. For future implications, this experiment could allow for model building and techniques in data mining in engineering education. The areas to improve will be the evaluation of factors associated amongst resource demands in airport operation. The performance assessment created PBL instructional designs of course objectives (e.g., integration of research to practices in engineering education). The empirical study unfounded assumptions to advance learning resources used with information sharing for engineering development in PBL and logistic regression analysis.

References

[1] F. Tanasache, M. Sorella, S. Bonomi, R. Rapone, and D. Meacci, "Building an emulation environment for cyber security analyses of complex networked systems," Networking and Internet Architecture, 2018

[2] M. Khanzadeh, S. Chowdhury, M. Marufuzzaman, M. A. Tschopp, and L. Bian, "Porosity prediction: Supervised-learning of thermal history for

direct laser deposition," Journal of Manufacturing Systems, vol. 47, pp. 69–82, 2018

[3] H. Cheng, D. Feng, X. Shi, and C. Chen, "Data quality analysis and cleaning strategy for wireless sensor networks," Journal on Wireless Communication and Networking, 2018.

[4] A. Akram, S. Iqbal, T. Iqbal, E. Munir, H. Khan, and M. Shafi, "Finding rotten eggs: A review spam detection model using diverse feature sets," KSII Transactions on Internet and Information Systems, vol. 12(10), pp. 5120–5142, 2018

[5] A. Ciuhu, N. Caragea, and C. Alexandru, "Modelling the potential human capital on the labor market using logistic regression in R," Romanian National Institute of Statistics, vol. 65(4), pp. 141-152, 2017

[6] D. Dua, and C. Graff, "UCI machine learning repository." Irvine, CA: University of California, School of Information and Computer Science, 2021

[7] N. Feng, Z. Su, D. Li, C. Zheng, and M. Li, "Effects of review spam in a firm-initiated virtual brand community: Evidence from smartphone customers," Information & Management, vol. 55(8), 1061–1070, 2018