Student Performance Prediction from E-mail Assessments Using Tiny Neural Networks

Nikhil Yadav[1] and Kajal Srivastava[2]
[1]Division of Computer Science, Mathematics and Science, St. John's University, Queens, New York, USA
[2]Business Communication, Jaipuria Institute of Management, Lucknow, India
yadavn@stjohns.edu, kajal.srivastava@jaipuria.ac.in

Abstract - **Predicting student performance using e-mail assessments can help in early interventions to better assist students sooner, rather than later, in STEM courses. In this paper, we propose *CorC-Net,* a tiny artificial neural network (ANN) that operates on limited data comprised of features scored from student assessments based on writing e-mails. ANNs are typically built using large scale data sets to truly realize their full potential; however, tiny neural networks overcome this problem by utilizing smaller batches of data making them easier to train. COrC-Net uses scored e-mails for content, organization, and clarity and classifies how students will perform. Formative instructor feedback provided between the assessments implies that CorC-Net is a more logical fit to simulate the "learning" process when human reaction to feedback and corrective action is involved. This is true especially in sequential course assessment tasks. In this paper, we show that COrC-Net outperforms other multi-class classification algorithms like decision trees, support vector machines, Gaussian Naïve Bayes, and K-nearest neighbors. CorC-Net's success in classifying student performance shows great potential in courses where long-term temporal assessment data is not available.**

Index Terms – Computer Science, Machine learning, Artificial neural networks, Learning analytics, Predictive modeling.

INTRODUCTION

It is well established that content, organization, and clarity are the pillars of successful written communication. Writing style, grammar, relevancy, and focus of the written word are good indicators of the core competencies of STEM students. Industry also requires individuals to have all-rounded communication skills, as far as communication is concerned. In particular, professionals have to churn out e-mails on a daily basis. With the advent of e-mails, there has been a paradigm shift in the manner of communication in the workplace. What once used to be communicated through letters and memos has been replaced by Instant Messages (IM) and e-mails. Resultantly, e-mails have become one of the most preferred channels for communication [5].

According to estimates, an employee today spends almost 25% of his/her time on e-mails [5]. Research has also shown that the efficiency of employees drops by 53% when they are away from e-mails [4]. 56% of employees surveyed by *USA Today* [6] stated that e-mails increase their productivity. Today, an average employee spends 49 minutes, while top managers spend around four hours on a daily basis over e-mails [10]. It is obvious that effective and efficient usage of e-mails can have a major economic impact on an organization.

The focus on producing candidates in STEM degrees, who have the perfect blend of oral as well as written communication, with specific emphasis to e-mail writing cannot be undermined. In addition to having knowledge pertaining to a particular domain, the aim is to produce candidates who have the right skill set and carry the right attitude required for a specific job. These skills are transferred to the workplace [8][11][13].

Currently, there is an immense thrust on Competency Based Education (CBE) or Outcome Based Education (OBE). This is a systematic learning process which has been talked about by academicians for over three decades [3][15].The movement related to competency based education began in the 1960's in USA. The core idea of CBE/OBE is to scientifically measure the performance of students. In order to do so, for any given Assessment Task (AT), rubrics are designed and learning outcomes are identified. Each rubric carries particular descriptors. This implies that the learning of every student is mapped against set parameters. Education accreditation bodies favor mapping students through this technique since it aids in *"requiring units of higher education to document how they know that graduating students have obtained the necessary competencies that support their respective degrees."* [12]

It is worth noting that there is a definite difference between competencies and learning outcomes [2]. Learning outcomes are described by faculty and subject experts, while the mastery of a student on a specific skill set is his/her "competency." Simply put, the learning outcomes belong to the academician while competencies belong to the student. Therefore, competency refers to defined knowledge/skill-set that a student can demonstrate to prove the acquisition of the listed knowledge/skillset. Competency is proof of the "how" while an outcome refers to "what" is meant to be achieved.

It is imperative that the outcome should be specific, measurable, achievable, realistic and time bound. In addition to this, students must also have the ability to practice the newly learned KSAs (Knowledge, Skills and Attitude). While doing so, students must be provided with regular feedback that should be formative and summative, during and after practice sessions, simultaneously. A blend of feedback and practice are essential and hold a higher degree of importance when compared to delivery through lectures only [14].

As a result, the e-mail competencies of students pertaining to studies can be mapped to rubrics based on three parameters (i.e. content, clarity, and organization) along with descriptors. Evaluating these parameters using e-mail assessments at the start of the course, can help in identifying students who require guidance to make them aware and address these issues.

With these objectives clearly defined, it is obvious that machine learning based models can be used to classify and eventually predict student success using such assessments. Doing this can help in early intervention strategies between the instructor and student, thereby promoting more engagement and motivation. Student performance prediction has been studied extensively in the learning analytics community. As noted in [1], previous work on predicting student success can be divided into modeling approaches incorporating:

- Generalized Linear Models (GLMs) like linear SVMs and logistic regression using attributes extracted from student grades, forums etc. [16]
- Artificial neural networks using reduced features resulting in poor performance. [9][16]

These models generally predict a binary classifier (pass or fail). They also require a lot of training data to make decisions. In this paper, an ANN approach was chosen to build COrC-Net, a tiny neural network using content, organization, and clarity scores of email assessments. These metrics comprise the feature set vector upon which student performance in e-mail communication is predicted. COrC-Net is different from approaches used in the past [1] for student performance prediction, such as conventional bi-directional (Long Short Term Memory) LSTM recurrent neural networks (RNNs) in the following ways:

- Requires a Smaller data set for training: COrC-Net can be applied with data from a few observations available (100s rather than 1000s).
- Does not require time-stamped data: Unlike past work on student performance prediction, time is not a parameter in the design of COrC-Net. It is a simple predictive ANN using reduced configuration layers.
- Multi-class classification: Unlike prior work, which is based on a binary outcome (pass or fail), COrC-Net can classify one of five classes: excellent (E), very good (V), average (A), poor (P), unacceptable (U). This is more realistic in a diverse class setting with students displaying different skill levels.

The remainder of this paper is organized as follows: The *Methodology* section describes the tools and rubrics used for assessing e-mail communication and the nature of data collected; it also details the *COrC-Net* architecture after formulating the classification problem; the benchmark baseline models are listed for comparison purposes. The *Results* section focuses on the descriptions of the training and validation phases together with evaluation of the COrC-Net classification and its comparison with traditional multi-class classification techniques. The *Conclusion* section highlights key takeaways from building COrC-Net and the analysis phase. It also proposes a future path for this research.

METHODOLOGY

I. Instruments for Data Collection

For the purpose of building a model, assessments in a business communication course taught at a management institute were used to evaluate e-mail communication. 118 students were involved in the data collection phase. Individuals were scored by professional faculty on the CLO (course learning outcome) stating that *"students should be able to utilize writing tools to write with clarity and organization."* The associated PLOs (program learning outcomes) were evaluated against this. To evaluate improvement (if any) in the students, two Assessment Tasks (AT-1 and AT-2) were assigned to the same group of 118 students.

AT-1 involved a scenario where an e-mail is written to a vendor requesting re-shipment of a wrong car part, due to an online listing error on the vendor website. AT-2 involved asking the student to write an e-mail to a probable mentor asking for career guidance post-graduation, for a field of work they are interested in. The student was instructed to find out the main things they would like to know about the particular field of employment, reflecting consideration for the reader and demonstrating their own interest in that type of job. The evaluation for both ATs was formative. Even though the assessments were used to explore classification models from a cohort of business communication students, the proposed data serves as the ground truth capable of integrating similar results from STEM students for such assessments when available.

While the first AT was taken during the seventh session of the course followed by formative feedback, the second AT was taken in the twelfth session of the course. This was again followed by formative feedback. It was intended that students could take the feedback from AT-1 and use it to improve their performance in AT-2. The tasks were formulated keeping in mind the PLO which stated that students should be able to:

- Communicate effectively and display good inter-personal skills
- Produce written documents and oral presentations that communicate ideas and information effectively for the intended audience and purpose

The rubrics used for the two assessments is shown in Table I, together with the points for each assessment criteria. Data was collected and scored using the rubric to count for a total of 118 samples for AT-1 and 117 samples for AT-2 (total of 235 samples after excluding students who missed either of the ATs). Upon assessing ATs, the instructor followed the rubrics to assign scores in compliance with the five classes.

TABLE I
RUBRICS FOR E-MAIL WRITING ASSESSMENTS

Criteria	Below Expectation	Meets Expectation	Exceeds Expectation
Content (40 points)	Subject line is lengthy, boring and unclear. Does not give any clear idea of the message contents.	Somewhat clear but less impactful subject line that gives some idea of the message contents.	Brief, clear, interesting and well-formed subject line that accurately gives the idea of the message contents.
	Main idea is supported by few or no explanations or facts.	Some explanations or facts used to support the main idea.	Appropriate explanations or facts used to support the main idea.
	points: 0-15	*points: 16-30*	*points: 31-40*
Organization (30 points)	Lack of understanding of proper salutation, closing and signature.	Either salutation, closing or signature is incorrect or missing.	Proper use of salutation. Email contains complimentary closing and signature with all required items like name, title, company name, contact information.
	Email seems to be a collection of unrelated ideas.	In the body, most ideas are expressed in a clear and organized manner.	In the body, all ideas are expressed in a clear and organized manner.
	points: 0-10	*points: 11-20*	*points: 21-30*
Clarity (30 points)	Words and phrases do not create a formal tone. In fact, words and phrases create an informal and/or rude tone. Contractions, slang, and/or emoticons were excessive.	Most words and phrases are appropriate for creating a formal tone without being rude; rare use of contractions, slang or emoticons.	Words and phrases are appropriate for creating a formal tone without being rude; no contractions, slang or emoticons used.
	Several errors in spelling, punctuation, capitalization, incorrect use of grammar that make the understanding of the message very difficult.	Spelling, punctuation, capitalization and grammar are somewhat correct and pose some difficulty in clear understanding of the mail.	Spelling, punctuation, capitalization and grammar are mostly correct and does not affect the clear understanding of the mail.
	Presentation of content is barely reader-friendly	Presentation of content is somewhat reader-friendly	Presentation of content is mostly reader-friendly
	points: 0-10	*points: 11-20*	*points: 21-30*

II. Problem Formulation

We can formulate the student performance prediction problem as a multi-class classification supervised learning problem. We have two assessment events in sequence (AT-1 followed by AT-2). For each student email assessment, their feature vector is a weighted sequence of CorC indicators together with the performance for that assessment. For 98% of the students we have two such vectors, one for AT-1 and one for AT-2.

For any assessment *n*, the input vector can be broken down into an input feature set described as a sequence of real numbers (1)

$$input = (C_n, O_n, Cl_n) \tag{1}$$

Where C is Content, O is Organization, and Cl is Clarity.

The output is the predicted student performance from the neural network and is a member of the full set of performance possibilities, as described by (2):

$$\begin{gathered} output \in \{E, V, A, P, U\}, \\ Where\ E = Excellent, V = Very\ Good, = Average, \\ P = Poor, and\ U = Unacceptable \end{gathered} \tag{2}$$

To be useful for ANN application, each of the member variables of the input are normalized using min-max scaling to be within the range [0, 1] using (3):

$$\begin{gathered} X_{std} = \frac{X - X_{\min(axis=0)}}{X_{\max(axis=0)} - X_{\min(axis=0)}} \\ X_{scaled} = X_{std} * (\max - min) - min \\ Where\ min, \max = feature\ range \end{gathered} \tag{3}$$

The transformation is calculated according to (4)

$$\begin{gathered} X_{scaled} = scale * X + min - X_{\min(axis=0)} * scale \\ Where\ scale = \frac{(\max - \min)}{X_{\max(axis=0)} - X_{\min(axis=0)}} \end{gathered} \tag{4}$$

The performance (output) data is categorical. *One-hot encoding* is used to convert the performance into 5-bit *one-hot vectors.* We can express this for a bit at the i'th position of the performance vector O, if it matches the corresponding performance indicator obtained by a particular student for an assessment as shown in (5).

$$f_o(x_i) = \begin{Bmatrix} 1, & if\ x_i = o_i \\ 0, & otherwise \end{Bmatrix} \quad (5)$$

III. Data and Training

Both ATs were distributively scored by the instructor teaching the course. The scoring was performed based on the writing style, grammar, relevancy, and focus of communication. The rubric in Table I was complied with for this purpose. It is worth noting that there is little correlation between these features, i.e., an e-mail may be grammatically correct but lack relevance and clarity, and vice versa. The breakdown of the performance for the students in the entire data set of 235 samples is shown in Table II. Both assessments were used collectively without considering the order of completion for the training phase.

TABLE II
PERFORMANCE BREAKDOWN

Performance	Samples
Excellent (E)	47
Very Good (V)	20
Average (A)	71
Poor (P)	73
Unacceptable	24

IV. Baseline Models

In order to assess the gains by switching to a tiny neural network, a few standard multi-class classification techniques were used to create a baseline. The models selected include:

- Support Vector Machines (SVMs)
- Decision Trees
- K-nearest neighbors (KNN)
- Gaussian Naïve Bayes

The training of all these models on the small data set we established, did not take long. These multi-class modeling techniques suffer from one inherent flaw, there is little potential to use any feedback to simulate the "learning" process. COrC-Net does not have this shortcoming due to the use of non-stochastic weight adjustments as highlighted in the next section.

V. COrC-Net Architecture

The traditional machine learning models mentioned in the previous section do not consider the potential feedback and back propagation phase that ANNs naturally entail. This is an advantage when instructor feedback is involved.

To establish/select an efficient configuration for COrC-Net, all the 235 samples in the data set were used. First the hold out method, with 50% of the data for testing and 50% for training (this was without application of K-fold cross validation initially) was applied. The results of the experimentations, using different numbers of hidden layers and nodes are shown for three such configurations in Figure I. Model 1 and Model 2 had one and two hidden layers with 8 nodes each respectively. Model 3 had three hidden layers with 8 nodes each. As evidenced from Figure I, this 3-layer architecture suffered from the least validation loss and achieved the highest accuracy of the three tiny ANN models explored. The final COrC-Net architecture in Figure II was chosen based on this result.

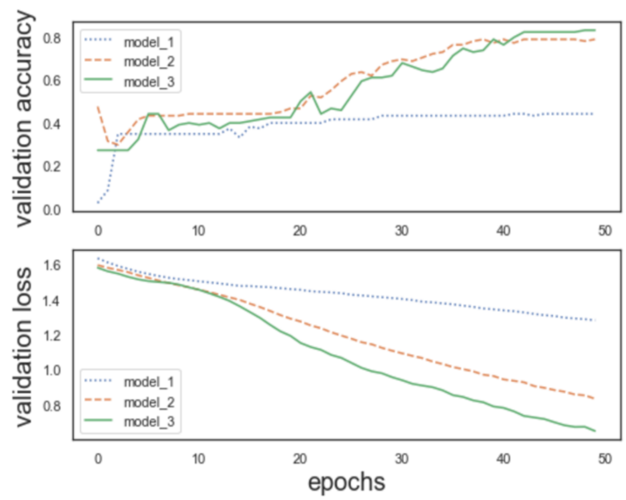

FIGURE I
ANN MODEL COMPARISONS

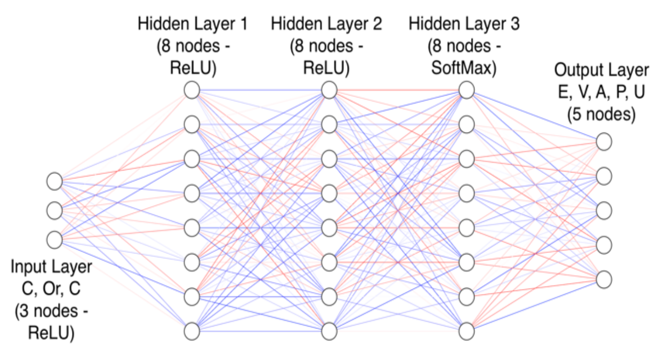

FIGURE II
CORC-NET ARCHITECTURE

This uses the ReLU (rectified linear unit) activation function into the hidden layers. The ReLU activation function converges fast, does not suffer from increased computation costs of the *sigmoid* and *tanh* functions, and is essentially non-linear.

For the inputs coming into the hidden layers, the ReLU function R(x) is applied and defined in (6).

$$R(x) = \begin{Bmatrix} 1, & for\ x < 0 \\ x, & for\ x \geq 0 \end{Bmatrix} \quad (6)$$

The Softmax function (7) is chosen to the output layer, since it is a generalized logistic activation function which suits our purpose of multi-class classification.

$$\alpha(Z)_j = \frac{e^{Z_j}}{\sum_{k=1}^{K} e^{Z_k}} \quad (7)$$

where Z is the vector of input values, K is the total number of input values, and Z_j is the *j'th* element in the vector of input values. It is worth noting that the dimension of time was not included in the data set that was used.

The *ADAM* optimization algorithm was used as the procedure to update network weights iteratively based on training data [7]. It is well documented that this method, compared to the classic stochastic gradient technique is more dynamic and maintains a learning rate for each network weight (parameter) which adapts as learning unfolds. For the purpose of COrC-Net, since human feedback was involved, this "adaptive" learning is a good simulator of actual response to corrective feedback.

RESULTS

In terms of training the models, ANN took an observably longer time to train compared to the other baseline models. The prediction accuracies for the 5 multi-class problem in the dataset were computed across all baseline models and subsequently compared with the accuracies generated using COrC-Net.

I. Baseline Results

For the baseline models, after 10-fold cross validation the mean validation accuracies and standard deviations computed are shown in Table III and the boxplot in Figure III.

TABLE III
BASELINE MODEL PERFORMANCE

Model	Mean Accuracy	Standard Deviation
SVM	0.788225	0.168977
DecTree	0.854529	0.096798
KNN	0.884601	0.090887
GaussianNB	0.879891	0.145197

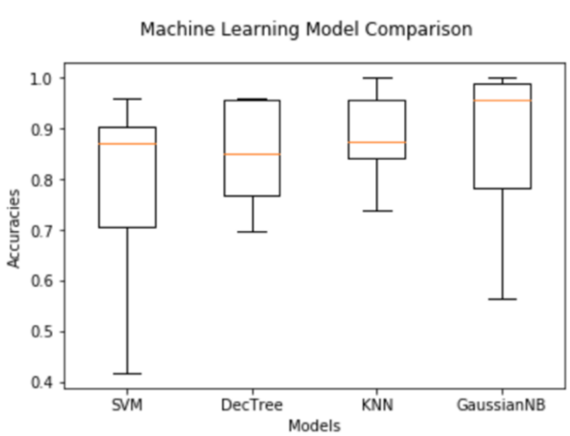

FIGURE III
BASELINE MODEL PERFORMANCES

Gaussian Naïve Bayes, Decision Trees, and the K-nearest neighbors supervised learning algorithms all performed in the mid to high 80%'s in accurately predicting the COrC data final classifier. The SVM algorithm did not perform as well. This may be due to some inaccuracies in labelling of support vectors due to the default choice of kernel that was used in the implementation.

II. COrC-Net

To evaluate COrC-Net's accuracy and to avoid any discrepancies due to differences in class frequencies (though marginal), a Receiver Operating Characteristic (ROC) was used for evaluating the quality of COrC-Net's prediction using the static holdout method. An ROC curve was created by plotting the true positive rate (TPR) against the false positive rate (FPR). The ROC curve obtained for COrC-Net is shown in Figure IV. The accuracy using this method was **94.8%** which was exceptional. However, comparisons with the benchmark need to be made on a level playing field using the same cross validation training-testing folds as the baseline methods.

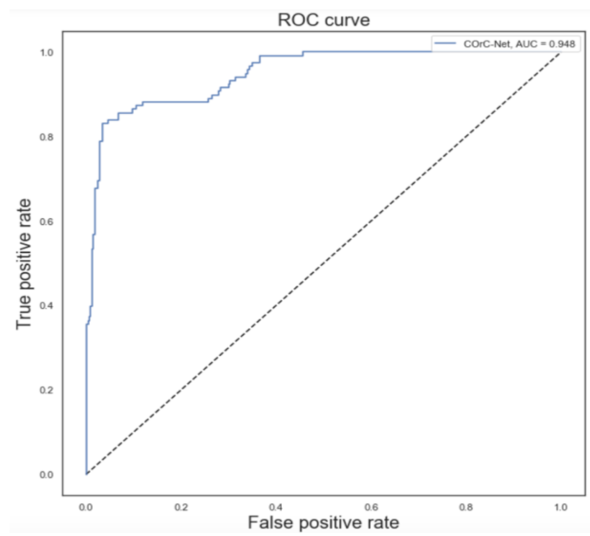

FIGURE IV
CORC-NET ROC CURVE (HOLDOUT METHOD)

To be able to confidently say that COrC-Net performs better than the baseline models discussed earlier, a similar 10-fold cross validation training-testing phase (as performed in the baseline models) was carried out. The mean accuracy obtained was **92%** with a standard deviation of **7%.** This is better than *all* the standard baseline models used as benchmarks in this paper. The limited network configuration to deal with small scale data proves that there is merit in using "tiny" implementations of ANNs.

Furthermore, the high classification accuracies imply that students facing difficulty in the class (i.e. students classified as *poor* or *unacceptable*) can be easily identified after feeding their assessment score into the COrC-net model. Instructors can intervene early to assist such students and address their communication deficiencies.

Conclusion

In this paper, the usefulness of tiny ANNs to predict student performances with limited data is highlighted. To our knowledge, the use of tiny neural networks to predict multi-class student performance has not been fully exploited. By fine tuning and building tiny neural networks to work on highly customized, reduced feature sets extracted from courses, this work shows that ANNs are better than KNN, SVM, Gaussian Naive Bayes and decision tree models for predicting student performance. They are also more suited theoretically to mimic the "learning" from feedback phase in sequential college course assessments.

Using pre-processed bespoke data from a business communication course makes these models more efficient as demonstrated in this paper. In the future, natural language processing (NLP) techniques can be devised to score email assessments and compare them to hand assigned instructor scores. COrC-Net's use in being able to make temporal predictions based on progressive assessments is also a future step, albeit for smaller STEM course assessment data sets with handpicked features.

References

[1] Byung-Hak, K., Vizitei, E., and Ganapathi, V., 2018. "Gritnet: Student performance prediction with deep learning." CoRR, 74–101.

[2] Cumming, A., and Ross, M. 2007. "The tuning project for medicine learning outcomes for undergraduate medical education in Europe." Medical Teacher 29:636–641.

[3] Greenhill, M., Metz, C., and Stander, P. 1982. "Performance-based public administration: A viable approach to adult learning." Teaching Political Science: A Journal of Social Sciences 9(4):197–204.

[4] Guffey, M. E. 2006. Business Communication: Process and Product. (5th Ed.). Mason, OH: Thomson Higher Education.

[5] Guffey, M. E. 2007. Essentials of business communications (7th ed.). Mason, OH: Thomson Higher Education.

[6] Hynes, G. E. 2008. Managerial communication: Strategies and applications. (4th ed.). New York: McGraw-Hill/Irwin.

[7] Kingma, D. P., and Ba, J. L. 2015. "Adam: A method for stochastic optimization." ICLR, San Diego.

[8] Kraiger, K. 2003. "Perspectives on training and development." In W. C. Borman, D. R. I., and Klimoski, R. J., eds. Handbook of psychology: Industrial and organizational psychology, volume 12. Hoboken, NJ: John Wiley & Sons.171–192.

[9] Fei, M. and Yeung, D., "Temporal Models for Predicting Student Dropout in Massive Open Online Courses," 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, 2015, pp. 256-263.

[10] O'Rourke, J. S., 2004. Management communication: A case analysis approach (2nd ed.). Upper Saddle River, NJ: Pearson-Prentice Hall.

[11] Prince, C., and Salas, E. 1993. "Training research for team-work in the military aircrew." In E. L. Wiener, B. G. K., and Helmreich, R. L., eds., Cockpit resource management. San Diego, CA: Academic Press. 337–366.

[12] Rivenbank, W., and Jacobson, W. S. 2014. "Three principles of competency-based learning: Mission, mission, mission." Journal of Public Affairs Education 20(2): 181–192.

[13] Salas, E., and Cannon-Bowers, J. A. 2001. "The science oftraining: A decade of progress." Annual Review of Psychology 52:471–499.

[14] Salas, E.; Tannenbaum, S.; Kraiger, K. et al. 2012. "The science of training and development inorganizations: What matters in practice." Psychological Science in the Public Interest 13:74–101.

[15] Voorhees, R. A. 2001. "Competency-based learning models: A necessary future." New Directions for Institutional Research 10(100):5–13.

[16] Whitehill, J.; Mohan, K.; Seaton, D. et al. 2017. "Delving deeper into mooc student dropout prediction." arXiv preprint (arXiv:1702.06404): 256–263