PREDICTING ADMISSION COUNSELING TRIUMPH OF COLLEGES USING NEURAL NETWORKS

*MAITREI KOHLI , **PRITI PURI
*Dept. Of Computer Science & Engg, **Dept. Of Information Technology
Lingaya’s institute of management & technology
Village Nachauli, old Faridabad-jasana road , Faridabad 121002, Haryana
INDIA

Abstract: Forecasting student admission in colleges and as a result ranking the colleges has intrigued many scholars and academic industry leaders as a difficult and challenging problem. In this study, the use of neural networks in forecasting the performance of technical colleges in the counseling session before it actually takes place is explored. We convert this forecasting problem into a classification one, i.e., we classify a college based upon its total number of seats filled in one of the five categories, ranging from ‘flop’ to ‘top’. Because our neural network model can predict the success of a college in gaining maximum admissions in an academic year before the actual admissions take place, it can be used as a powerful decision aid by university management, colleges and anticipating students. Since we have used the evaluation version of the software, we are therefore getting the maximum accuracy of 72%. This is mainly due to the reason that the evaluation version does not incorporate all the features required to compute accurate result. Employing the neural network model by using the 10-fold cross validation methodology shows that the neural networks do a much better job of predicting in this setting.

Key-Words: - Forecasting, neural networks, independent variables, MLP architecture, Placements, Fee – structure, Faculty

1 Introduction
Forecasting student admission in colleges and as a result ranking the colleges has intrigued many scholars and academic industry leaders as a difficult and challenging problem. For most analysts, admission counseling is a process of premonition and wild guess, due largely to the difficulty and uncertainty associated with predicting the college demand. Such unpredictability makes the college business a risky endeavor for the investors to take in today’s highly competitive world. No one can tell us how a college is going to fare in its exam i.e. the counseling session. Not until the actual event takes place.

Despite the difficulty associated with the unpredictable nature of problem domain, several researchers have attempted to develop models for forecasting the success of colleges in gaining maximum admissions in an academic year, primarily using statistics based forecasting approach. Most of them have tried to predict the total performance of all colleges after the actual counseling is held. However the results obtained were not very accurate for decision support.

In our study, we explore the use of neural networks in forecasting the performance of technical colleges in the counseling session before it actually takes place. We convert this forecasting problem into a classification one, i.e., we classify a college based upon its total number of seats filled in one of the five categories, ranging from ‘flop’ to ‘top’.

The remainder of this paper is organized as follows. Section 2 briefly reviews the basics of ANN & also presents the comparison between the statistical techniques & our ANN approach. Section 3 gives the details of our methodology.
by specifically talking about the data, the neural network model, the experiment methodology and the performance measures used in this study. Next, the experimental results of the neural network model are shown and explained. Finally, the Section 5 of the paper discusses the overall contribution of this study, along with its limitations and further research directions.

2 Review of ANN
A neural network's ability to perform computation is based on the hope that we can reproduce some of the flexibility and power of the human brain by artificial means. Basically, a neural network is a machine that is designed to model the way in which the brain performs a particular task or function of interest. The network is usually implemented by using electronic components or is simulated in software on a digital computer.

The property of the neural network that is of primary significance is the ability of the network to learn from its environment, & to improve its performance through learning. A neural network learns about its environment through an interactive process of adjustments applied to its synaptic weights & bias level. After completing the learning process successfully the network is ready to be deployed for independent functioning.

Neural networks have been applied to an increasing number of real-world problems of considerable complexity. The most important advantage is that artificial neural networks are capable of solving problems that are too complex for conventional technologies – problems that do not have an algorithmic solution or that solution is too complex to be found.

The application domain of neural networks today touches almost the entire sphere of science. These include: Association, Clustering, Classification, Pattern Completion, Regression & Generalization, Forecasting, Optimization etc to name a few.

Many application bibliographies exist. However, none of these include an application in forecasting the performance of colleges in gaining max admission during the counseling session. This study is one of the first to attempt the use of neural networks for addressing this challenging problem that has drawn the attention of many researchers in such areas of decision support systems and management science.

We differentiate our study from the others as follows. First, there is no reported study on using neural networks to predict the college performance in counseling session. Our study seems to be the first attempt of its kind in this problem domain. Another distinguishing feature of our study comes from its longitudinal nature. Our study is based on 10 consecutive years of data that covers movies released between 1997 and 2007. Our study also compares the difference between those of individual years and the combined data set of all 10 years. The results suggest that our neural networks model performs better than the one reported in the literature.

3 Methodology & data & variable definition
Though commonly known as black box approach or heuristic method, in the last decade, artificial neural networks have been studied by statisticians in order to understand their prediction power from a statistical perspective. These studies indicate that there are a large number of theoretical commonalities between the traditional statistical methods, such as discriminant analysis, logistic regression, and multiple linear regressions, and their counterparts in neural networks, such as multi-layered perceptron, recurrent networks, and associative memory networks.

In our study, we explore the use of neural networks in forecasting the college selections, by students going for technical education before the actual admissions take place. Here, we convert the forecasting problem into a classification problem. Multi layer perceptron (MLP) neural network architecture is known to
be a strong function approximator for prediction & classification problems.[13]. MLP is capable of learning arbitrarily complex non-linear functions to an arbitrary accuracy level. Thus it is a candidate for exploring the rather difficult problem of mapping college performance to the underlying characteristics. The error back-propagation algorithm presents the best mapping; it is thus used in this approach.

3.1 Data & Variable Definitions
In our study, n technical colleges, affiliated to the same university & located in a state in northern India are being used. The variable of interest in the study is the total number of seats filled during the counseling sessions based on merit only. It does not include any management seat, NRI seat etc. We have converted this forecasting problem into a classification one i.e. a college based on it’s total number of seats filled (on the basis of merit) is classified in one of the five categories, ranging from ‘FLOP’ to ‘TOP’. We plan to use eight different types of independent variables. Each categorical-independent variable is converted into 1-of-N binary representations. Thus we get a number of pseudo- representations that increases the independent variable count from 8 to 28.[1].

A neural network treats these pseudo variables as different mutually exclusive information channels. All pseudo representations of a categorical variable will be given a value of 0, except the one that holds true for the current case, which will be given the value of 1. For e.g. the variable ‘Fee Structure’ would be represented with two pseudo variables – High & Low. Now as an example consider the college IIT, it’s a govt. college & hence the fee structure is very nominal, therefore because of its low fees the value of variable ‘LOW’ would be set to 1 & that of ‘high’ would be set to 0.

We now present a brief description of the variables chosen.

3.1.1 Placement: - The most common thing mostly people see before choosing any technical college is its placement. The better is its placement the more famous a college would be. Two main things contribute to the worth of the placement parameter- one is the number of students placed in different companies & secondly the category (standard) of companies coming for placement. If big & renounced companies are coming for campus placements then the market value of that college would definitely increase. There are 4 possible ratings in this category, viz, High (placement>80%), Fair (60 – 80%), Average (40-60%) & Poor (<40%).
3.1.2 Academics: - Each affiliated college competes with all other colleges for the same bunch of students anticipating to take admission. The crucial feature that differentiates these colleges is their academic standard. Obviously the college offering better study environment, or in other words has good academics, has an edge over others. We classify this independent variable into 4 categories- Excellent, Good, Fair, Average.

3.1.3 Location: - Is yet another very important feature that determines a college’s fate. Any technical college located at some prime location, e.g., within a city will surely be benefited, as all students might want to take admission into a college that is closer to home. Thus we categorize Location onto 3 main categories. This are- Within a city, NCR, Distant area.

3.1.4 Facilities Provided: - Any student seeking admission in a technical college looks forward to get some basic facilities from the college. These common facilities that everyone wants to get are: 1) Transportation Facility 2) Good Hostel 3) Canteen or Mess 4) Co-Curricular activities. Based on these four parameters we classify this independent variable in 4 classes- Good; Fair; Average & Poor.

3.1.5 Faculty: - The faculty of a college plays a very vital role in deciding the future if the college. Good, dedicated & dutiful faculty ensures good academics, which in turn ensures good results & therefore better placements. The decision variable Faculty may be divided into 3 main classes- Good, Fair, Average.

3.1.6 Fee structure: - This forms yet another important parameter in deciding any college’s fate. Three different situations reside within this variable.

a) The college is good in all respects but the fee is nominal, mostly in case of govt. colleges. Then such a college is surely going to earn maximum number of students.

b) The college is good in all respects but the fee structure is also high, like in private colleges, in such cases people, people don’t mind taking admission, despite high fee structure because they’re getting all the facilities & features desired from a good technical college. So the parameter high fee structure then takes a backseat.

c) The last situation is that the college is not good i.e. it does not have good infrastructure, good faculty & does not provide any other facilities also but its fees is Low. Now in such a situation despite the fact that the fees are low, people will decide against taking admission in such colleges. This implies that Low fee parameter cannot save a college if it is poor in other areas. We’ve assigned 2 values to this variable. One value is ‘Low’ & the other value is ‘High’

3.1.7 Infrastructure: - what are the sure shot ingredients of a technical college - The Labs, The Library (preferably digital), Seminar Halls, and Lecture rooms Computer center etc. These techno savvy features make a college fit for survival & these constitute the infrastructure of a college. We thus classify this independent variable into 5 classes, viz, Excellent, Good, Fair, Average & Poor.

3.1.8 Market Value: - This decision variable, represented by 3 pseudo representations – High, Medium & Low, shows the market value of the college in current scenario. It tries to depict how much the college is able to earn through its name. Data for this variable has been collected by surveys i.e. by considering expert opinion as well as common man’s notion.

A summary of above-mentioned variables is given in Table 1. In all there are 28 decision variables & 5 output variables that we intend to use in our approach.
Table 1: Summary of independent Variables

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Independent Variable Name</th>
<th>No. Of Values</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1)</td>
<td>Placement</td>
<td>4</td>
<td>High (&gt;80%), Fair (60-80%), Average (40-60%), Poor (&lt;40%)</td>
</tr>
<tr>
<td>2)</td>
<td>Academics</td>
<td>4</td>
<td>Excellent; Good; Fair; Average</td>
</tr>
<tr>
<td>3)</td>
<td>Location</td>
<td>3</td>
<td>City; NCR; Distant place</td>
</tr>
<tr>
<td>4)</td>
<td>Facilities Provided</td>
<td>4</td>
<td>Good; Fair; Average; Poor</td>
</tr>
<tr>
<td>5)</td>
<td>Faculty</td>
<td>3</td>
<td>Good; Fair; Average</td>
</tr>
<tr>
<td>6)</td>
<td>Fee Structure</td>
<td>2</td>
<td>High; Low</td>
</tr>
<tr>
<td>7)</td>
<td>Infrastructure</td>
<td>5</td>
<td>Excellent; Good; Fair; Average; Poor</td>
</tr>
<tr>
<td>8)</td>
<td>Market Value</td>
<td>3</td>
<td>High; Medium; Low</td>
</tr>
</tbody>
</table>

Table 1: Summary of independent Variables

3.2 Experiment methodology

Many researchers in the past few years have studied the performance of neural networks in predicting a variety of classification problems over a wide range of different business settings. Many of these studies, however, were based on a single experiment, and/or the method of selecting the training and testing samples was not clear. We believe that because of the stochastic nature of the neural network training, better experimental design methods are necessary to develop the objective performance measures of neural networks.

As opposed to using a single neural network experiment to base our results upon, we chose to follow a more statistically sound experimental design methodology, called k-fold cross-validation. In k-fold cross-validation, also called rotation estimation, the complete dataset (D) is randomly split into k mutually exclusive subsets (the folds: D1, D2,.., Dk) of approximately equal size. The classification model is trained and tested k times. Each time (t2 {1,2,.., k}), it is trained on all but one folds (D\Dt) and tested on the remaining single fold (Dt). The cross-validation estimate of the overall accuracy is calculated as simply the average of the k individual accuracy measures.[6]. Since the cross-validation accuracy would depend on the random assignment of the individual cases into k distinct folds, a common practice is to stratify the folds themselves. In stratified k-fold cross-validation, the folds are created in a way that they contain approximately the same proportion of predictor labels as the original dataset. Empirical studies showed that stratified cross-validation tend to generate comparison results with lower bias and lower variance when compared to regular k-fold cross-validation [6]. In this study, to estimate the performance of neural network classifier a stratified 10-fold cross-validation approach is used. In 10-fold cross-validation, the entire data set is divided into 10 mutually exclusive subsets (or folds) with approximately the same class distribution as the original data set (stratified). Each fold is used once to test the performance of the classifier that is generated from the combined data of the remaining nine folds, leading to 10 independent performance estimates.

4 Results

Our approach of using Neural Networks in this field aims to classify the technical colleges in Five Output categories & thus forecasts their performance in the counseling session even before the actual event happens. We HAVE IMPLEMENTED this project using an evaluation version of the software NeuroSolutions5.[7]. Since we have used the
evaluation version of the software, we are therefore getting the maximum accuracy of 72%. This is mainly due to the reason that the evaluation version does not incorporate all the features required to compute accurate result. This research project is the first attempt to use Neural Networks for addressing this challenging problem that combines two different application domains of Forecasting & Classification & brings out the much-desired output. This model would be highly beneficial to the:

a) University Management- They’ll be able to rank their colleges. Also the over all standard of the university can be improved by identifying week colleges & thus by taking corrective actions.

b) College Management- They would be benefited for they can come to know their strong & weak points & can therefore make appropriate changes.

c) Anticipating Students- and finally the students seeking admission would be benefited for they’ll have all the comparative information that is needed & therefore at a glance they can decide which college to opt for.

5 Conclusion

Compared to the other model types (i.e. logistic regression, discriminant analysis and classification and regression trees), using the exact same experimental conditions, as reported in Section 4, neural networks performed significantly better. Because the neural network models built as part of this study are designed to predict the success of a college in gaining maximum admissions in an academic year before the actual admissions take place, they can be used as a powerful decision aid by university management, colleges and anticipating students. In any case, the results of this study are very intriguing, and once again, prove the continued value of neural networks in addressing difficult prediction problems.

Beyond the accuracy of our results in predicting admission-counseling success, these neural network models could also be adapted to forecast the examination results of the various colleges. The particular parameters used within the model could be altered using the already trained neural network model, in order to better understand the implications of different parameters on the end results-examination results. During this alternative experimentation process, the college administration could find out, with a fairly high accuracy level, how the students are going to perform in their exams.

Just like many other stochastic modeling techniques, neural networks start from a random set of weights. By utilizing the architectural parameters such as learning algorithm, learning rate, number of PEs in the hidden layers, etc., it adjusts those weights to create a map between the input and the output vectors. Correct choice of those architectural parameters plays a great role in developing better neural network models.

From an application perspective, once developed to a production system, such a neural network model can be made available (via a web server or as an application service provider) to academic decision makers, where individual users can plug in their own academic related parameters to forecast the potential success of a college in the counseling session. A neural network model can be designed in a way such that it can calibrate its weights (continuous self learning) by taking into account new samples as they become available. Much additional work, in terms of modeling extensions, further experimentation for testing the performance, and applications to other media product demand forecasting, remains to be done.

References:


[14] Predicting box-office success of motion pictures with neural networks ,Ramesh Sharda*, Dursun Delen Department of Management Science and Information Systems, William S. Spears School of Business, Oklahoma State University, Stillwater, OK 74078, USA