

A Data Mining Approach to Predict the Performance of College Faculty

J. Jelsteen, K. Anandan

^{1,2}Assistant Professor, Nehru College of Management, Coimbatore, Tamil Nadu, India

ABSTRACT

The success of a given job calculated against preset known standards of accuracy, completeness, cost, and speed. Performance is deemed the fulfillment of a commitment, in a manner that releases the performer from all liabilities under the agreement. In this paper we summarize our research by compare the Bayesian network classifiers for prediction of faculty performance, which helps in overall growth of the college. The data mining approach used for extracting useful models from the institutional database is able to extract certain anonymous trends in faculty performance when assessed across several parameters.

Keywords : Data Mining Techniques, Bayesian Networks Classifiers, Classification, Prediction

I. INTRODUCTION

Performance evaluation is a constructive process to acknowledge the performance of a non-probationary career employee. An employee's evaluation shall be sufficiently specific to inform and guide the employee in the performance of her/his duties. Assessment as a dynamic process produces data, which acts as performance indicator for an individual and subsequently impacts on the decision making of the stakeholders as well as individual. The main objective of educational institutes is to provide quality of education to its students and to improve the quality of institutions.

Basically there are two approaches of data analysis that can be used for extracting models describing significant classes or to predict future data trends. These two forms are classification and prediction.

Classification techniques are supervised learning techniques that classify data item into predefined class label. It is one of the most useful techniques in data mining to build classification models from an input data set. The used classification techniques commonly build models that are used to predict future data trends.

Prediction models predict continuous valued functions. Predictive analytics encompasses a variety of statistical techniques from predictive modelling, machine learning, and data mining that analyze current and historical facts to make predictions about future or otherwise unknown events.

II. LITERATURE SURVEY

Data mining is the process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses, or other information repositories [1]. Data mining refers to extracting knowledge from large amount of data.

According to Romero et al., [2], there are increasing research interests using data mining in education. Aranuwa and Sellapan [3] used directed modeling that is an intelligent technique for evaluation of instructors' performance in higher institutions of learning, and proposed an optimal algorithm and designing a system framework suitable for predicting instructors' performance as well as recommended necessary action to be taken to aid school administrators in decision making considering the limitations of the classical methodologies.

Business organizations are interested to settle plans for correctly selecting proper employees. After recruiting employees, the management becomes concerned about the performance of these employees where they build evaluation systems in an attempt to preserve the good performance of employees [5].

Used a Naive Bayes classifier to predict job performance in a call center with the aim of knowing what levels of the attributes are indicative of individuals who perform well. By using operational records, they predicted future performance of sales agents, achieving satisfactory results [6].

The main focus of this paper is to predict the faculty performance by using bayesian network classifiers. In this paper, the attributes such as faculty age, qualification, experience, publications, funded projects, students feedback, attendance and extension activities are used to examine the results in order to predict the faculty performance at the institution level.

III. DATA MINING PROCESS

Data mining refers to extracting or “mining” knowledge from large amount of data. Data mining as a synonym for another popularly used term, knowledge discovery from data or “KDD” The goal of this technique is to find pattern that was previously unknown data [7]. A historical overview Data Mining and its future directions in terms of standard for a Knowledge Discovery and Data Mining process model is given in [4].

The steps of knowledge discovery process as discussed as follows,

Selection : Selecting data relevant to the analysis task from the database.

Preprocessing : Removing noise and inconsistent data, combining multiple data sources.

Transformation:
Transforming data into appropriate forms to perform data mining.

Data Mining : Choosing a data mining algorithm which is appropriate to pattern in the data, extracting data patterns.

Interpretation/Evaluation

Interpreting the patterns into knowledge by removing redundant or irrelevant patterns. Translating the useful patterns into terms that human understandable.

Classification & Prediction

Classification models predict categorical class labels; and prediction models predict continuous valued functions.

Building the Classifier or Model

- This step is the learning step or the learning phase.
- In this step the classification algorithms build the classifier.
- The classifier is built from the training set made up of database tuples and their associated class labels.
- Each tuple that constitutes the training set is referred to as a category or class. These tuples can also be referred to as sample, object or data points.

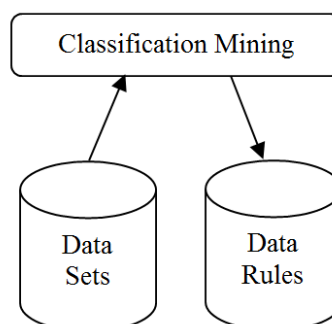


Figure 1. Architecture

What is classification?

Classification is the process of using a model to predict unknown values (output variables), using a number of known values (input variables). For example we might want to predict whether a stock market is currently a bull or a bear market, based on a number of market indicators, or we might want to predict whether a patient has a certain disease given a number of symptoms.

In order to perform classification, first we need to model the relationship between the input variables and

the output variables we are predicting. This process involves learning a model using data in which both the input variables and the output variables are present. Expert opinion can also be used to build/enhance a model. This model can subsequently be used on unseen data in which only the input data is present, in order to predict the output variables.

Classification is termed a supervised learning approach, because a model is trained specifically for predicting the output variable.

Typically, the term classification is concerned with predicting discrete variables. The term regression is used when predicting continuous variables.

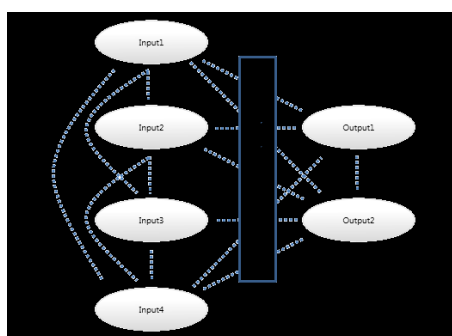


Figure 2. Model

Classification with Bayesian networks

Bayesian networks are widely used to perform classification tasks, with the following advantages.

- Based on probability theory
- Allows rich structure
- Can mix expert opinion and data to build models
- Backwards reasoning - in addition to predicting outputs given inputs, we can use output values to infer inputs
- Support for missing data during learning and classification
- Support latent variables for modeling hidden relationships
- Support time series classification

Image 1 depicts the possible structure of a Bayesian network used for classification. The dotted lines denote potential links, and the blue box is used to indicate that additional nodes and links can be added to the model, usually between the input and output nodes.

Using Classifier for Classification

In this step, the classifier is used for classification. Here the test data is used to estimate the accuracy of classification rules. The classification rules can be applied to the new data tuples if the accuracy is considered acceptable.

Comparison of Classification and Prediction Methods

Here are the criteria for comparing the methods of Classification and Prediction.

Accuracy

Accuracy of classifier refers to the ability of classifier. It predict the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

Speed

This refers to the computational cost in generating and using the classifier or predictor.

Robustness

It refers to the ability of classifier or predictor to make correct predictions from given noisy data.

Scalability

Scalability refers to the ability to construct the classifier or predictor efficiently; given large amount of data.

Interpretability

It refers to what extent the classifier or predictor understands.

II. METHODS AND MATERIAL

1. Methodology

The main objective of the proposed methodology is to build the classification model. It consists of the following two-step processes of data classification.

- Training Data
- Testing Data

In the first step, a model that describes a predetermined set of class is built by analyzing a set of training dataset. Each dataset is assumed to belong to a predefined class. In the second step the model is tested using a different data set that is used to estimate the classification accuracy of the model. There are several techniques that can be used for classification such as decision tree, Bayesian methods and so on.

In this paper, we have used Bayesian Network Classifier to build the classification model for the purpose of predicting the faculty performance. This prediction is performed on the basis on various attributes. The following table describes the faculty dataset. The data collected randomly from different engineering colleges at Coimbatore district.

Variable	Variable Type
Faculty ID	Numerical
Faculty Name	Categorical
Gender	Categorical
Qualification	Categorical
Experience	Numerical
Attendance	Categorical
Individual subject pass percentage	Numerical
Student Feedback	Numerical
Journal publications	Numerical
Books publications	Numerical
Seminar/Conference Organized	Numerical
Seminar/Conference Attended	Numerical
Acted as a resource person at other institute	Numerical
Government funded project	Numerical
Extension activities	Categorical

Table 1: Dataset Description

How to apply the Gaussian to the Bayes Classifier?

The application here is very intuitive. We assume the Density Estimation follows a Gaussian distribution. Then the prior and the likelihood can be calculated through the Gaussian PDF. The critical thing here is to identify the Gaussian distribution (i.e. find the mean and variance of the Gaussian). The following 5 steps are a general model to initialize the Gaussian distribution to fit our input dataset.

1. Choose a probability estimator form (Gaussian)
2. Choose an initial set of parameters for the estimator (Gaussian mean and variance)
3. Given parameters, compute posterior estimates for hidden variable
4. Given posterior estimates, find distributional parameters that maximize expectation (mean) of joint density for data and hidden variable (Guarantee to also maximize improvement of likelihood)
5. Assess goodness of fit (i.e. log likelihood) If not stopping criterion, return to (3).

From $P(S | M) = P(S)$
the rules of probability imply:
 $P(\sim S | M) = P(\sim S)$
 $P(M | S) = P(M)$
 $P(M \wedge S) = P(M) P(S)$
 $P(\sim M \wedge S) = P(\sim M) P(S)$
 $P(M \wedge \sim S) = P(M) P(\sim S)$
 $P(\sim M \wedge \sim S) = P(\sim M) P(\sim S)$

“The sunshine levels do not depend on and do not influence who is teaching.” can be specified very simply :

$$P(S | M) = P(S)$$

“Two events A and B are statistically independent if the probability of A is the same value when B occurs, when B does not occur or when nothing is known about the occurrence of B”.

III. RESULTS AND DISCUSSION

The performance of the faculty members are shown here using two variable output

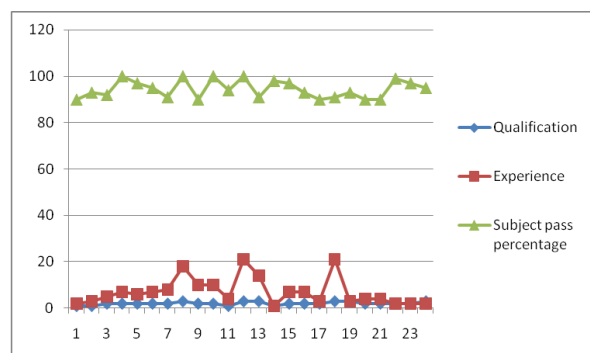


Figure 3. Academic Pass Percentage

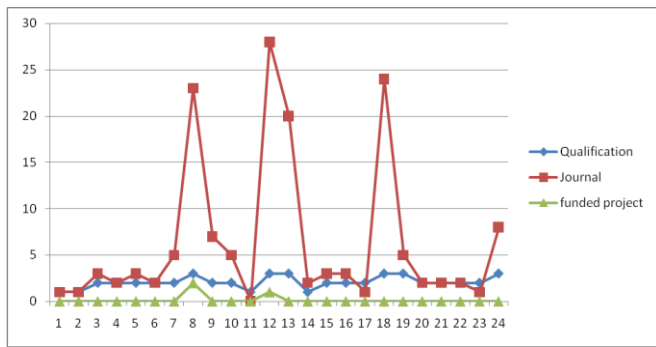


Figure 4. Qualificatin Vs Funded Project & Journal Publication

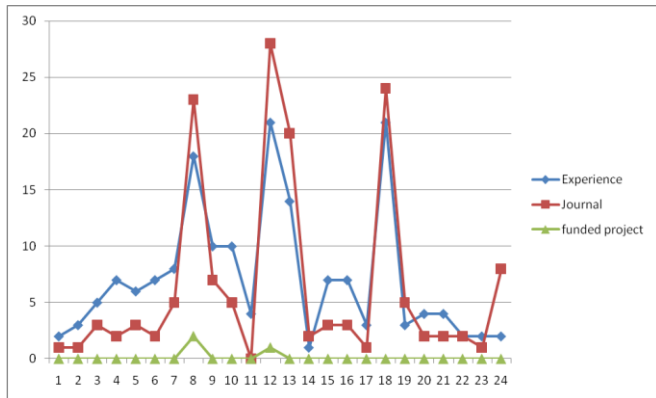


Figure 5. Experience Vs Funded Project & Journal Publication

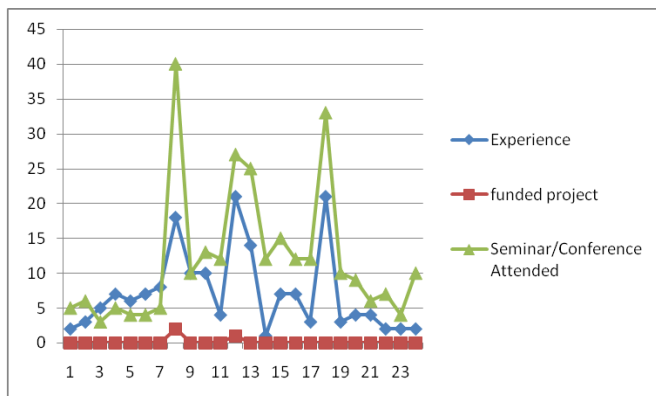


Figure 6. Experience Vs Funded Project & Seminar Participation

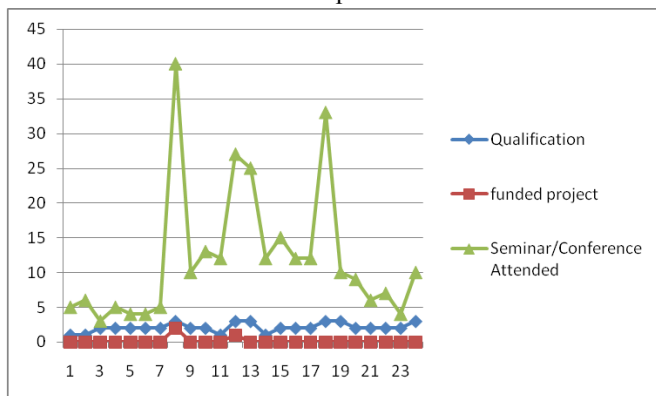


Figure 7. Qualification Vs Funded Project & Seminar Participation

Based on the experimental result we found the following information

- The academic result (pass percentage) point of view, minimum experience and qualified faculty were produced good results.
- Highly qualified and experienced faculty members were produced more journal publications and received funded projects.

Highly qualified and experienced faculty members were attended /organized conferences and seminars, so that they achieved in their funding project.

IV.CONCLUSION

This paper focused on the possibility of building a classification model for predicting faculty performance. In overall, institute as a whole can perform better by improving its faculty. By applying data mining algorithm of Bayesian Networks Classifiers, the institute administration will be able to make groups of faculty members with different parameters for future use. The performance and efficiency of this research can be improved by increasing the performance parameters like research centre, faculty exchange programme at national and international level.

V. REFERENCES

- Ogunde A.O and Ajibade D.A (2014): A Data Mining System for Predicting University Students' Graduation Grades Using ID3 Decision Tree Algorithm. Journal of Computer Science and Information Technology. March 2014, Vol. 2, No 1, pp 21 – 46.
- Romero C., Ventura S., Garcia E. (2008) Data mining in course management systems: Moodle case study and tutorial, Computers & Education, Vol. 51, No. 1, pp. 368-384, 2008
- Aranuwa F.O., and Sellapan P.,(2013): A data mining model for evaluation ofinstructors'performance in higher institutions of learning usingmachine learning algorithms, International Journal of Conceptions on Computing and Information Technology Vol. 1, sue 2, Dec' 2013; ISSN: 2345 – 9808
- Kurgan, L.A., Musilek, P. (2006). A survey of knowledge discovery and Data Mining Models, The Knowledge Engineering Review, 21(1), pp 1 - 24

- [5]. hein, C.F., Chen, L.F., (2008). Data Mining to improve personnel selection and enhance human capital: A case study in high technology industry, *Expert Systems with Applications*, 34(1), pp 280–290
- [6]. Valle, M.A., Varas, S., Ruz, G.A., (2012). Job performance prediction in a call center using a Naive Bayes classifier, *Expert Systems with Applications*, 39(11), pp 9939–9945
- [7]. Han and Kamber, “Data Mining: Concepts and Techniques”, Second Morgan Kaufman Publisher, 2006.