ISSN: 2321-7782 (Online)

Volume 4, Issue 1, January 2016

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Money Ball for Mining Faculty Academic Performance using Naïve Byes Classifier

Dhiran R. Karale[1]
Department of computer science and engineering,
Jhulelal Institute Of Technology, Lonara,
Nagpur, India

Pushpa Chutel[2]
Department of computer science and engineering,
Jhulelal Institute Of Technology, Lonara,
Nagpur, India

Abstract: The focus of this paper is to depict the vast landscape of literature related to enterprise performance measurement in a concise and comprehensible manner for researchers and practitioners. We focus particularly on the enterprise as the unit of analysis and consider measurement systems from stakeholders at all levels. A broad range of considerations will be explored, ranging from micro-level considerations such as employee performance measurement to macro-level considerations such as enterprise measurement systems. Moreover, we discuss measurement-related problems identified in practice and solutions proposed in academic literature. To illustrate this evolution of measurement knowledge over time, we discuss the effects of metrics from three distinct viewpoints: selecting the right metrics, creating and implementing measurement frameworks; and metrics for decision making.

Key words: money ball approach to academia, Data sources, proposed methodology, proposed plan work, survey on paper.

I. INTRODUCTION

First we explore the idea of selecting the right metrics. In order to develop a common grounding, we expand on the concept of measurement and fundamental problems individuals and organizations face regarding measurement. This discussion focuses around common mistakes and metric selection methodologies, considering respective implications on individual behavior. We provide an example from professional baseball to demonstrate how thinking creatively can ensure metrics correspond to value added activities and increase human productivity. Second, we describe the creation and implementation of measurement frameworks. Attributes of macro-level frameworks such as Kaplan and Norton’s Balanced Scorecard will be compared with other complementary approaches. We also discuss the many classifications of these frameworks, from “traditional” to “contemporary” systems, considering “structural” and “procedural” models, understanding temporal aspects, and identifying unique challenges and benefits from a case study of a micro (bottom-up) measurement system implementation. Third, we discuss the role of metrics in decision making. In particular, we consider how to use metrics with imperfect information. To supplement various academic viewpoints provided, we offer a practical discussion regarding guidance for decision makers for focusing on the right problem and dealing with imperfect information – contextually relevant for managers. This paper is an introductory guide for both practitioners and researchers to gain a better understanding about enterprise performance measurement. This guide is not intended to be collectively exhaustive, but indeed makes a point to articulate readings relevant to each section that one can consult for further information. In considering metrics selection, implementation and decision making there will never be a silver bullet – “a single development, in either technology or management technique, which by itself promises even one order of magnitude improvement in productivity, in reliability, in simplicity” .The practical implications of all three metric subjects are highly dependent on a variety of factors, to include but not limited to: the maturity of an organization and their processes; top down or bottom-up measurement system implementation; the industry being considered; the unit of analysis, such as people or business units; and the perspective taken during measurement. The principles and conclusions discussed in this paper will be depicted universally such that they can be applied in any context. For a brief

© 2016, IJARCSMS All Rights Reserved

overview of the literature discussed which provides a rough correlation regarding how select representative readings fit into each subtopic.

II. Money Ball Approach to Academia

In a recent study, the approach is presented a data-based method to predict the top-cited papers by using data available at the time of publication or at while that published. Their result shows that the citation network centrality of a paper is a better predictor of the future impact of the paper. It is possible to extend this thought to predict the academic or curriculum impact of a professor or faculty. Many metrics can be used to rate a professor such as citations, publications, student satisfaction, student feedback and regard in their field of study. Metrics such as citations and publications are readily available from open sources online including Google Scholar. In this study, we first investigate the distribution of salary and then Fellowship of professional societies regarding different explanatory metrics of a professor or a faculty. Using data mining concept we make the following contributions in this study: First we examine the correlation between salaries and citations, publications second, we expand our study to relate these metrics with the more formal ways professors are recognized such as the fellowship of associations such as the IEEE or ACM. We develop the classifiers to support the idea of data-based approach to academic decision-making process. In this paper we use the semi supervised classifier for a better result or for better use.

III. Literature Survey

Money ball is the approach which has been applied to many sections and variety of studies had been done on this topic. We have gone through some of the research done in this section. In the emerging field of legal informatics, lawyers and law firms integrate data and analytics into the quantitative legal predictions in litigation and transactions In business, a study conducted by Brynjolfsson et al. in 2011 showed that firms that adopt data-driven decision making strategies are observed to have output and productivity 5-6% higher than what would be expected given the firms' other information Dimitris et al. presented a data-based method to predict the top-cited papers by using data available at the time of publication. Their result shows that the citation network centrality of a paper is a good predictor of the future impact of the paper.

1] Performance Evaluation of Classification Methods for Heart Disease Dataset

In medical domain, a vast number of researches are being carried out due to the fact that it is inherently complex application domain. For instance, data mining techniques have been applied on breast cancer (Seera and Lim, 2014), colon cancer (Antonelli *et al.*, 2012), heart disease (Shouman, *et al.*, 2011), (Shouman, *et al.*, 2012), diabetes (Antonelli *et al.*, 2013), (Baralis *et al.*, 2010), (Mahoto, *et al.*, 2014), and eye patients (Mahoto, *et al.*, 2014). Especially several data mining techniques are exploited, for instance, classification technique (Shouman, *et al.*, 2011), sequential pattern mining (Baralis *et al.*, 2010), (Mahoto, *et al.*, 2014), clustering (Antonelli *et al.*, 2013), (Mahoto, *et al.*, 2014). Classification technique, particularly, is also widely used in medical domain. For example, improving Naïve Bayes classifier's accuracy (Abraham, *et al.*, 2006), hybrid intelligent model (Seera and Lim, 2014), and hybrid classification approach (HCA) (Chen, *et al.*, 2014) for medical dataset. The study in (Soni *et al.*, 2011) presents data mining techniques for heart disease prediction. In particular, decision tree, k-NN, neural networks and Bayesian classification are discussed. The study in (Shouman, Turner and Stocker, 2012) applies k-NN for heart disease dataset and showed its improved results.

2] The Prediction of Students' Academic Performance Using Classification Data Mining Techniques.

The data were collected from 8 year period intakes from July 2006/2007 until July 2013/2014 that contains the students' demographics, previous academic records, and family background information.

Decision Tree, Naïve Bayes, and Rule Based classification techniques are applied to the students' data in order to produce the best students' academic performance prediction model.

3] Automated Classification of Web Sites using Naive Bayesian Algorithm.

The proper classification has made these directories popular among the web users. The exponential growth of the web has made it difficult to manage human edited subject based web directories. The World Wide Web (WWW) lacks a comprehensive web site directory. Web site classification using machine learning techniques is therefore an emerging possibility to automatically maintain directory services for the web. This paper describes Naïve Bayesian (NB) approach for the automatic classification of web sites based on content of home pages.

4] Is Naïve Bayes a Good Classifier for Document Classification?

Document classification is a growing interest in the research of text mining. Correctly identifying the documents into particular category is still presenting challenge because of large and vast amount of features in the dataset. In regards to the existing classifying approaches, Naïve Bayes is potentially good at serving as a document classification model due to its simplicity. The aim of this paper is to highlight the performance of employing Naïve Bayes in document classification. Results show that Naïve Bayes is the best classifiers against several common classifiers (such as decision tree, neural network, and support vector machines) in term of accuracy and computational efficiency.

IV. DATA SOURCES

In this paper we collect detailed information from three publicly available data sources or from open sources. The detail about salary statistics is collected from an online database. The number of publications and citations, as well as the number of years in the research field . The last data source is the IEEE and ACM official websites, which provide us a full list of IEEE and ACM Fellows in which they shows faculty publication details means how many times he published the paper or what type of implementation he doneor what is his activity in curriculum. We classified the professors into two categories: being a Fellow of either association (or both), Once the data was scraped and collected, some basic analytics were done to obtain a better insight into the data. The professors can also be grouped based on their designation to better see the distribution:

Designation	Count	Average
Distinguished Professor	91	$124,317
Research Professor	23	$85,553
Associate Professor	4306	$62,049
Assistant Professor	3549	$53,311

Fig 1. SALARY DISTRIBUTION FOR GROUPED PROFESSORS BASED ON DIFFERENT DESIGNATION

V. PROPOSED RESEARCH METHODOLOGY

We are proposing data centric analysis of faculty performance where we consider many publication data like Google scholar and IEEE to get the citation about the faculty. Apart from this we are also considering salary data with publications data and checking whether are related to each other. After all this, we are building a classifier which will take few records of data and build training set and then rest of the records are passed to classifier to check whether they can get fellowship of IEEE/ACM.to improve and manage the organization to be more efficient to mange the faculty activities.

VI. PROPOSED PLAN WORK

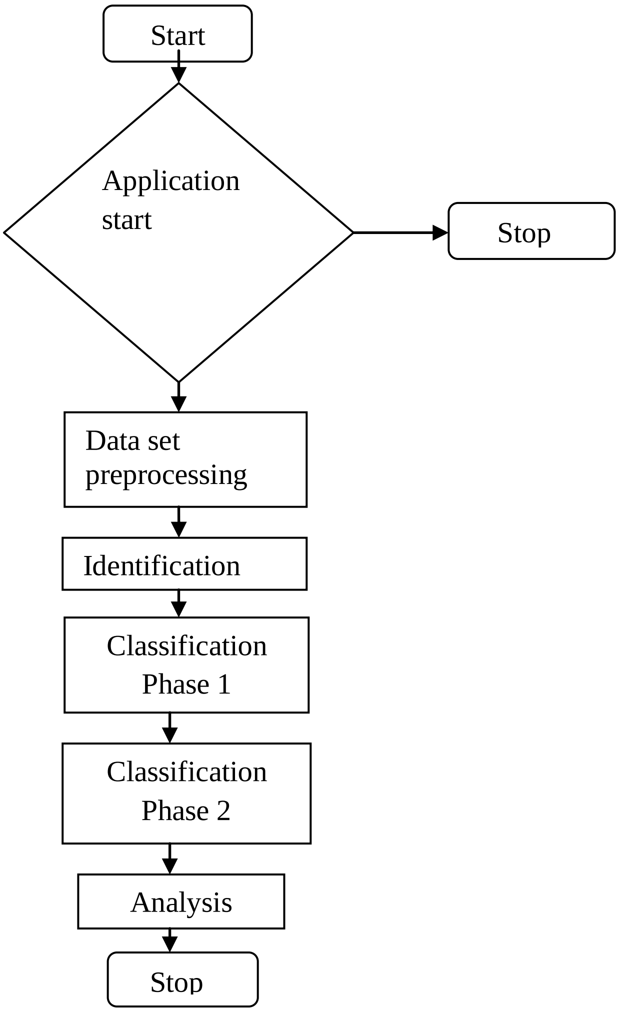
Module 1: Basic UI generation, framework building and dataset gathering

Module 2: In this phase we will preprocess the dataset. This step will include cleaning of dataset, and identifying each field and insert them in relational database.

Module 3: Classification will be performed in this phase using naïve byes classifier. First training set will be used to build the classifier and then classification steps will be performed on test set.

Module 4: Analysis and testing will be performed in this module. This will also include the comparison of proposed approach with the existing one.

VII. Data Flow Diagram



VIII. Technique / Algorithm

The classification technique used in the research is two tree-based ensemble classification algorithms, Adaboost and Random Forest. But as we can see from the dataset it is high dimensional data and we should use naïve byes classifier for better result. This is the place where we can improve the result.

IX. Expected Outcome

The input to the system will be the dataset from publication house and university records of the faculty. This data will be pre-processed and will be passed to classifier to get the result. After successful completion of the experiment, we are expecting the faculty performance measurement and improvement based on data mining techniques.

References

1. Lewis, Michael. Money ball: The art of winning an unfair game. WW Norton & Company, 2004.
2. Money ball for Lawyers: How Data and Analytics are Transforming the Practice of Law. Available from: https://lexmachina.com/wpcontent/uploads/2013/08/Moneyball-for-Lawyers-Owen-Byrd-The-Bottom-Line-April-2013.pdf
3. Money ball For Music: The Rise of Next Big Sound. Available from: http://www.forbes.com/sites/zackomalleygreenburg/2013/02/13/moneyball-for-music-the-rise-of-next-big-sound/
4. Erik Brynjolfsson, Lorin M. Hitt, Heekyung Kim: Strength in Numbers: How does data-driven decision-making affect firm performance? ICIS 2011. https://www.researchgate.net/.../228221..

© 2016, IJARCSMS All Rights Reserved

5. Bertsimas, Dimitris, et al.: Money ball for Academics: Network Analysis for Predicting Research Impact. Available At SSRN 2374581, 2014. https://github.com/.../blob/.../proposal.t..

6. Yiming Liu, Dong Xu, Ivor W. Tsang, Jiebo Luo: Textual Query of Personal Photos Facilitated by Large-Scale Web Data. IEEE Trans. Pattern Analysis and Machine Intelligence 33(5): 1022-1036, 2011.

7. Data Mining: Introduction Why Mine Data?http://staffwww.itn.liu.se/~aidvi/courses/06/dm/lectures/lec1.pdf

8. The Prediction of Students' Academic Performance Using Classification Data Mining Techniques http://www.m-hikari.com/ams/ams-2015/ams-129-130-2015/p/ahmadAMS129-130-2015-2.

9. DBertsimas,E Brynjolfsson , S Reichman ….- Available at SSRN….2014 – papers.SSRN.com

10. Moneyball for Academia: Toward Measuring and Maximizing Faculty Performance and Impact A Nocka, D Zheng, T Hu, J Luo - Data Mining Workshop (…, 2014 - ieeexplore.ieee.org B. Badur and S. Mardikyan, "Analyzing Teaching Performance of Instructors Using Data Mining Techniques," Informatics in Education, vol. 10, no. 2, pp. 245–257, 2011.