

# **Employee Performance Prediction Model**

*Submitted in partial fulfillment of the requirements for the degree of*

**Master of Philosophy**

*by*

**Rupali Sunil Wagh**

**(Roll No.0935012)**

**Supervisor:**

**Prof Joy Paulose**



**Department of Computer Science  
CHRIST UNIVERSITY BANGALORE  
2010**

# Declaration

I hereby declare that the dissertation entitled '**Employee Performance Prediction Model**' submitted for the M.Phil. Degree is my original work and the dissertation has not formed the basis for the award of any degree, associateship fellowship or any other similar titles.

Place: Bangalore

Date:

Rupali Sunil Wagh

Roll No.: 0935012

Signature of the student

## **Certificate**

This is to certify that the dissertation entitled '**Employee Performance Prediction Model**' is a bonafide research work carried out by Rupali Sunil Wagh, student of M.Phil. (Computer Science) Christ university, Bangalore, during the year 2009-2011, in partial fulfillment of the requirements for the award of the Degree of Master of philosophy and that the dissertation has not formed the basis for the award previously of any degree, diploma, associateship, fellowship or any other similar title.

Place: Bangalore

Date:

Prof. Joy Paulose  
Signature of the Guide

# Approval Sheet

This thesis entitled **Employee Performance Prediction Model** by **Rupali Sunil Wagh** is approved for the degree of Master of Philosophy in Computer Science.

Examiners:

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_

Supervisor (s)

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_

Chairman:

\_\_\_\_\_

(Seal)

Date: \_\_\_\_\_

Place: \_\_\_\_\_

## **Abstract**

With the dominance of knowledge power in the success of an organization, competent human resource has become crucial for realization of organizational objectives. Human Resource Management, HRM is a set of tasks to maintain and develop a proficient human resource. A performance appraisal process helps the HRM in identifying the strengths and weaknesses of an employee. This evaluation of employee is based on several different parameters according to the work domain and organizational objectives. This activity of employee evaluation has a high significance in making strategic decisions of manpower planning than just salary reviews. The objective of the prediction model constructed in the study is to assist HR personnel in decision making by predicting the performance of an employee.

This study, has developed an Employee Performance Prediction Model which can predict the performance category of an employee. Academic industry data is used for the analysis. The supervised learning approach of classification has been first applied to the performance data to construct the model. The methodology adapted by the study is a five step process of Data preprocessing, Dimensionality Reduction, Application of classification algorithm to build the classifier and Evaluation of the classifier. Based on the hierarchical characteristics of the data, three different approaches – Direct, Hierarchical and Integrated are followed for the construction of the classifiers. A comparative analysis of the classifiers' performances based on prediction accuracy and error rate is performed. Unsupervised learning approach is used to understand natural similarities among the employees to enhance the evaluation process.

The thesis is organized into five chapters. Chapter1, Introduction, starts with a brief overview of HRM and presents the objective of the study and the problem statement. It further describes a typical KDD process and role of data mining in a KDD process. Chapter 2, Literature review, discusses background work in the field of data mining and HRM. Different models and techniques used in the context are described. Chapter 3, Methodology elaborately discusses the processes and techniques used for constructing the Employee Performance Prediction Model. Starting with the block diagram, the chapter elaborates the sequence of steps that led to the results. Chapter 4, Results and Discussion

discusses the performance statistics of the different classifiers constructed in the study. Chapter 5, Conclusion, summarizes the inferences concluded based on the results obtained. The chapter also discusses limitations and challenges and concludes with future scope of the study.

**Keywords: Human Resource Management, Performance Appraisal, Knowledge discovery, Data Mining, Supervised Learning, Unsupervised Learning, Classification and Prediction**

<b>Table of Contents</b>	
<b>Topics</b>	<b>Page No</b>
<b>Abstract</b>	<b>V</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>Abbreviations</b>	<b>ix</b>
<b>1. Introduction</b>	<b>1</b>
<b>1.1. HRM – An overview</b>	<b>1</b>
<b>1.2. Objective of the research</b>	<b>6</b>
<b>1.3. Problem Statement</b>	<b>6</b>
<b>1.4. Data Mining in Knowledge Discovery Process</b>	<b>7</b>
<b>1.4.1. Knowledge Discovery Process</b>	<b>8</b>
<b>1.4.2. Data Mining – Knowledge Acquisition Tool</b>	<b>10</b>
<b>1.5. Thesis Organization</b>	<b>12</b>
<b>2. Literature Review</b>	<b>13</b>
<b>2.1 Back ground Work</b>	<b>13</b>
<b>2.1.1 A generic DM Model for HRM</b>	<b>14</b>
<b>2.1.2 Talent Forecasting Model For HRM</b>	<b>15</b>
<b>2.1.3 Performance Prediction – Data warehouse Approach</b>	<b>16</b>
<b>2.1.4 Work Behavior Prediction – Rough Set Approach</b>	<b>17</b>
<b>2.2 Review of Related Techniques and Methods</b>	<b>17</b>
<b>2.2.1 Data pre processing</b>	<b>17</b>
<b>2.2.2 Dimensionality Reduction</b>	<b>18</b>
<b>2.3 Review of Supervised and Unsupervised Learning Methods</b>	<b>19</b>
<b>2.3.1 Supervised Learning Method – Classification</b>	<b>19</b>
<b>2.3.2 Unsupervised Learning Methods</b>	<b>22</b>

<b>3</b>	<b>Methodology</b>	<b>24</b>
	<b>3.1 Performance Prediction Model – Block Diagram</b>	<b>24</b>
	<b>3.2 Data set description</b>	<b>25</b>
	<b>3.2.1 Data Characteristics</b>	<b>25</b>
	<b>3.2.2 Multi-Level Data Representation</b>	<b>27</b>
	<b>3.3 Data Understanding and Preprocessing</b>	<b>28</b>
	<b>3.3.1 Correlation Analysis</b>	<b>29</b>
	<b>3.3.2 Integration of Data Sets</b>	<b>30</b>
	<b>3.3.3 Filling of Missing Values</b>	<b>30</b>
	<b>3.3.4 Data Generalization</b>	<b>30</b>
	<b>3.3.5 Representation of descriptive Attributes</b>	<b>31</b>
	<b>3.4 Supervised Learning approach</b>	<b>31</b>
	<b>3.4.1 Elementary Prediction model</b>	<b>32</b>
	<b>3.4.2 Refined Classification Model</b>	<b>37</b>
	<b>3.4.2.1 Hierarchical Approach</b>	<b>37</b>
	<b>3.4.2.2 Direct Approach</b>	<b>41</b>
	<b>3.4.2.3 Integrated Approach</b>	<b>42</b>
	<b>3.5 Unsupervised Learning Approach</b>	<b>42</b>
	<b>3.6 Classification Algorithms</b>	<b>45</b>
	<b>3.6.1 Decision Tree Classifiers</b>	<b>45</b>
	<b>3.6.2 Naïve Bayes Classifier</b>	<b>46</b>
	<b>3.6.3 Information Gain – Attribute evaluation Measure</b>	<b>47</b>
	<b>3.7 Tools Used</b>	<b>48</b>
<b>4</b>	<b>Results and Discussion</b>	<b>50</b>
	<b>4.1 Classifier Evaluation</b>	<b>50</b>
	<b>4.1.1 Elementary Classifier Evaluation</b>	<b>51</b>
	<b>4.1.2 Intermediate Classifier</b>	<b>61</b>
	<b>4.1.3 Hierarchical Classifier Evaluation</b>	<b>61</b>
	<b>4.1.4 Direct Classifier Evaluation</b>	<b>63</b>
	<b>4.1.5 Integrated Classifier Evaluation</b>	<b>66</b>



<b>4.2 Cluster Analysis</b>	<b>67</b>
<b>4.3 Discussion Of Results</b>	<b>69</b>
<b>5 Conclusion</b>	<b>72</b>
<b>5.1 Summary</b>	<b>72</b>
<b>5.2 Challenges and Limitations</b>	<b>73</b>
<b>5.3 Future enhancement</b>	<b>75</b>
<b>Bibliography</b>	
<b>Acknowledgments</b>	

## List of Figures

Figure 1:	Functions of HRM	3
Figure 2:	Sample Appraisal Form	5
Figure 3:	Knowledge Discovery Process	9
Figure 4:	DM for Talent Management	15
Figure 5:	Attribute Subset selection	18
Figure 6:	Steps of Supervised learning method	20
Figure 7:	Block Diagram of the Model	24
Figure 8:	Two level Structure of appraisal Data	27
Figure 9:	Hierarchical Representation of Parameters	28
Figure 10:	Elementary Classification Model	32
Figure 11:	Hierarchical Classification approach	39
Figure 12:	Direct Classifier Approach	41
Figure 13:	Integrated Approach for attribute Selection	42
Figure 14:	ROC curve for 4_M	54
Figure 15:	ROC curve for Hierarchical Classifier	63
<b>Figure 16</b>	ROC curve for Direct Classifier	65
Figure 16:	ROC curve for Integrated Classifier	67
Figure 17:	Silhouette Plot	68
Figure18:	Assignment Of objects to clusters	69

## **List of Tables**

<b>Table 1</b>	<b>Performance parameters – D4</b>	<b>53</b>
<b>Table 2</b>	<b>Performance parameters – D4_2</b>	<b>55</b>
<b>Table 3</b>	<b>Performance parameters – D4_3</b>	<b>57</b>
<b>Table 4</b>	<b>Performance parameters – D4_4</b>	<b>59</b>
<b>Table 5</b>	<b>Performance parameters – D4_5</b>	<b>60</b>
<b>Table 6</b>	<b>Performance parameters – D4_6</b>	<b>61</b>
<b>Table 7</b>	<b>Performance parameters – Hierarchical Classifier</b>	<b>62</b>
<b>Table 8</b>	<b>Performance parameters – Direct Classifier</b>	<b>64</b>
<b>Table 9</b>	<b>Performance parameters – Integrated Classifier</b>	<b>66</b>

## **Abbreviations**

Human Resource Management (HRM)

Knowledge Discovery (KDD)

Data Mining (DM)

Knowledge, Skills and Abilities (KSA)

Key Performance Indicator (KPI)

True Positive (TP)

False Positive (FP)

Receiver Operating Characteristics (ROC)

## Table of contents

Topics	PageNo.
<b>Abstract</b> .....	<b>iv</b>
<b>List of figures</b> .....	<b>vii</b>
<b>List of tables</b> .....	<b>viii</b>
<b>Abbreviations</b> ....	<b>ix</b>
<b>1. INTRODUCTION</b> .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
1.1 IT IN PHARMACEUTICAL DOMAIN.....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
1.2 OBJECTIVE OF THE RESEARCH .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
1.3 PROBLEM STATEMENT .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
1.4 AGILE METHODOLOGY AN OVERVIEW .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
1.4 COMPONENT BASED SOFTWARE DEVELOPMENT AN OVERVIEW ...	<b>ERROR! BOOKMARK NOT DEFINED.</b>
<b>2. LITERATURE REVIEW</b> .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
2.1 PRESENT SCENARIO OF SOFTWARE DEVELOPMENT IN PHARMACEUTICAL COMPANY	<b>ERROR! BOOKMARK NOT DEFINED.</b>
2.2 AGILE METHODOLOGIES /CBD ADOPTED IN THE DOMAIN.....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
<b>3. AGILE METHODOLOGY</b> .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
3.1 EXTREME PROGRAMMING.....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
3.2 ADAPTIVE SOFTWARE DEVELOPMENT.....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
3.3 DYNAMIC SYSTEMS DEVELOPMENT METHOD .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
3.4 SCRUM.....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
3.5 CRYSTAL .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
3.6 FEATURE DRIVEN DEVELOPMENT.....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
3.7 AGILE MODELING.....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
4. COMPONENT BASED SOFTWARE DEVELOPMENT.....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
4.1 PURPOSE OF COMPONENT-BASED SOFTWARE DEVELOPMENT .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
4.1.1 HISTORY .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
4.1.2 BENEFITS .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
4.1.3 MARKETPLACE .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
4.2 REQUIREMENTS FOR SOFTWARE RE-USE .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
4.3. DIFFICULTIES ASSOCIATED WITH SOFTWARE RE-USE .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
<b>5. PHARMACEUTICAL ANALYSIS</b> .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
5.1 MEDICINE MANUFACTURING .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
5.1.1 DRUG RESEARCH .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
5.1.3 BULK ACTIVE OPERATION: .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
<b>6. IMPLEMENTATION</b> .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
6.1 SCENARIO DEPICTING THE USAGE OF CBD .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
6.2.1 EXISTING SYSTEM.....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
CONFIGURATION AND SETUP .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
TECHNICAL ARCHITECTURE .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
6.2 SCENARIO DEPICTING USAGE OF AGILE METHODOLOGY .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
<b>COMPARISON OF AGILE AND COMPONENT BASED SOFTWARE DEVELOPMENT.</b>	<b>ERROR! BOOKMARK NOT DEFINED.</b>
BEHAVIORAL FEASIBILITY .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>

**7. QUESTIONNAIRE .....ERROR! BOOKMARK NOT DEFINED.**  
**8 RESULTS AND DISCUSSION .....ERROR! BOOKMARK NOT DEFINED.**  
**9.CONCLUSION .....ERROR! BOOKMARK NOT DEFINED.**  
**APPENDIX .....ERROR! BOOKMARK NOT DEFINED.**  
**QUESTIONNAIRE RESULTS FOR ALL SAMPLES .....ERROR! BOOKMARK NOT DEFINED.**  
**LITERATURE CITED .....ERROR! BOOKMARK NOT DEFINED.**  
**PUBLICATIONS.....ERROR! BOOKMARK NOT DEFINED.**