

**MOBILE DATA MINING USING ENHANCES
CLASSIFICATION TECHNIQUE TO IMPROVE
CORRECTLLY CLASSIFIED INSTANCES**

Dissertation submitted in fulfilment of the requirements for the Degree of

MASTER OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

SATBIR KAUR

41400004

Supervisor

HARJEET KAUR



School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab (India)

Month JUNE Year 2017

@ Copyright LOVELY PROFESSIONAL UNIVERSITY, Punjab (INDIA)

Month, Year

ALL RIGHTS RESERVED



TOPIC APPROVAL PERFORMA

School of Computer Science and Engineering

Program : 1792::M. Tech- CSE(Computer Science and Engineering)(Part Time)

COURSE CODE : CSEP545 REGULAR/BACKLOG : Regular GROUP NUMBER : CSERGD0271

Supervisor Name : Harjeet Kaur UID : 12427 Designation : Assistant Professor

Qualification : _____ Research Experience : _____

SR.NO.	NAME OF STUDENT	REGISTRATION NO	BATCH	SECTION	CONTACT NUMBER
1	Satbir Kaur	41400004	2014	K1418	09779403201

SPECIALIZATION AREA : Programming-I Supervisor Signature: _____

PROPOSED TOPIC : Assessment of the services of mobile environment on data mining with decision trees

Qualitative Assessment of Proposed Topic by PAC		
Sr.No.	Parameter	Rating (out of 10)
1	Project Novelty: Potential of the project to create new knowledge	7.00
2	Project Feasibility: Project can be timely carried out in-house with low-cost and available resources in the University by the students.	7.00
3	Project Academic Inputs: Project topic is relevant and makes extensive use of academic inputs in UG program and serves as a culminating effort for core study area of the degree program.	6.67
4	Project Supervision: Project supervisor's is technically competent to guide students, resolve any issues, and impart necessary skills.	7.50
5	Social Applicability: Project work intends to solve a practical problem.	6.17
6	Future Scope: Project has potential to become basis of future research work, publication or patent.	6.33

PAC Committee Members		
PAC Member 1 Name: Janpreet Singh	UID: 11266	Recommended (Y/N): Yes
PAC Member 2 Name: Harjeet Kaur	UID: 12427	Recommended (Y/N): Yes
PAC Member 3 Name: Sawal Tandon	UID: 14770	Recommended (Y/N): Yes
PAC Member 4 Name: Vikas Verma	UID: 11361	Recommended (Y/N): Yes
PAC Member 5 Name: Dr. Ramandeep Singh	UID: 14105	Recommended (Y/N): Yes
DAA Nominee Name: Kanwar Preet Singh	UID: 15367	Recommended (Y/N): Yes

Final Topic Approved by PAC: Assessment of the services of mobile environment on data mining with decision trees(Needs clarity)

Overall Remarks: Approved (with major changes)

PAC CHAIRPERSON Name: 11011::Dr. Rajeev Sobti

Approval Date: 25 Oct 2016

ABSTRACT

There are several techniques that are used in data mining, each one having advantages but also disadvantages. To find out which one is most appropriate for our case, when we want to use our databases in a decision-maker process, it is good to know that “data mining doesn’t reduce the require to know your company, to recognize your data, or to recognize systematic method”. So, having information about our data business and data mining techniques we can decide what we will use. Or we can try them all (if we have enough time, money and data) and find out which one is the best in our case. DT is one of the significant study methods in categorization. It builds its best tree form by selecting significant association character. While collection of test quality and division of trial sets are two vital parts in construction trees. altered decision tree method will accept different technology to resolve these nuisance. conventional algorithms contain CART, SPRINT, ID3, SLIQ etc. ID3 is the depiction of decision tree method. It is simple to appreciate and has quick classify haste which is related to big datasets. Many DT algorithms are better base on it, like C45, CART. But these algorithms additional or less have some trouble in collection of test character, sort of sample, memory consumption of data and the prune of trees etc. Currently, researchers have current many improvements.

The dataset used in this research is based on mobile environment obtained from WirelessMon software. This report is based on the findings maximum used of mobile service. The results in this report are based on data from mobile service related.

As we seem at Data Mining tools, we notice that readily available are dissimilar algorithms use for creating a result creation (or analytical study) scheme. There are algorithms for creating result trees such as CHAID and C4.5 along with algorithms for formative identified nearest neighbor (KNN) or clustering when operational on classification. The objective of this research is to seem at one exacting decision tree algorithm called improved algorithm and how it can be use with data mining for mobile service. The reason is to influence huge amounts of data and change it into in sequence that can be used to create a decision.

DECLARATION STATEMENT

I hereby declare that the research work reported in the dissertation entitled " **MOBILE DATA MINING USING ENHANCES CLASSIFICATION TECHNIQUE TO IMPROVE CORRECTLLY CLASSIFIED INSTANCES**" in partial fulfilment of the requirement for the award of Degree for Master of Technology in Computer Science and Engineering at Lovely Professional University, Phagwara, Punjab is an authentic work carried out under supervision of my research supervisor Mrs.Harjeet Kaur I have not submitted this work elsewhere for any degree or diploma.

I understand that the work presented herewith is in direct compliance with Lovely Professional University's Policy on plagiarism, intellectual property rights, and highest standards of moral and ethical conduct. Therefore, to the best of my knowledge, the content of this dissertation represents authentic and honest research effort conducted, in its entirety, by me. I am fully responsible for the contents of my dissertation work.

Signature of Candidate

SATBIR KAUR

R.NO.41400004

SUPERVISOR’S CERTIFICATE

This is to certify that the work reported in the M.Tech Dissertation entitled “**MOBILE DATA MINING USING ENHANCED CLASSIFICATION TECHNIQUE TO IMPROVE CORRECTLY CLASSIFIED INSTANCES**”, submitted by **Satbir Kaur** at **Lovely Professional University, Phagwara, India** is a bonafide record of her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree.

Signature of Supervisor

Harjeet kaur

Date:

Counter Signed by:

1) Concerned HOD:

HoD’s Signature: _____

HoD Name: _____

Date: _____

2) Neutral Examiners:

External Examiner

Signature: _____

Name: _____

Affiliation: _____

Date: _____

Internal Examiner

Signature: _____

Name: _____

Date: _____

ACKNOWLEDGEMENT

I would like to express my gratitude and deep appreciation to my supervisor Ms. Harjeet kaur for her guidance and positive suggestion. I would also like to express my thanks to my friends, classmates and to the people who have supported in completing my Dissertation-II. Finally I would like to thank my Mom, Dad, Husband and my Brother for their love, support and patience over the semester. This Dissertation is dedicated to them.

TABLE OF CONTENTS

CONTENTS	PAGE NO.
CHAPTER1: INTRODUCTION	1
1.1. DATABASE	
1.2. DATA MINING TECHNIQUES	2
1.3. TYPES OF DATA	3
1.4. DIFFERENCE BETWEEN DATA WAREHOUSE AND DATABASES	4
1.5 DATA MINING	5
1.6 PROBLEMS FACED IN DATA MINING	6
1.7 DECISION TREES	7
CHAPTER2: REVIEW OF LITERATURE	9
CHAPTER3: PRESENT WORK	22
3.1 PROBLEM FORMULATION	22
3.2 SCOP OF THE STUDY	22
3.3 FACILITIES REQUIREMENT	23
3.4 OBJECTIVES OF THE STUDY	23
3.5 RESEARCH METHODOLOGY	23
CHAPTER4: RESULTS AND DISCUSSION	25
4.1 DATASET SCREENSHOTS	25
4.2 COMPARISON WITH EXISTING TECHNIQUE	35

CHAPTER5: CONCLUSION AND FUTURE SCOPE	38
5.1 CONCLUSION	38
5.2 FUTURE SCOPE	38
REFERENCES	39
APPENDIX	

LIST OF TABLES

TABLE NO.	TABLE DESCRIPTION	PAGE NO.
Table 4.1	Analysis between CART, C45 and IDTDMA algorithms to find out the accuracy in Correctly Classified Instances	36
Table 4.2	Analysis between CART, C45 and IDTDMA algorithms to find out the accuracy in Incorrect Classified Instances	37
Table 4.3	Analysis between CART, C45 and IDTDMA algorithms to find out the Error Rate	38

LIST OF FIGURES

FIGURE NO.	FIGURE DESCRIPTION	PAGE NO.
Figure1.1	Shows database management system	1
Figure1.2	Different techniques are Used in Data mining	3
Figure1.3	Show Data Mining Process	5
Figure1.4	Example of Decision Tree	8
Figure3.1	Flow chart of research methodology	24
Figure4.1	WEKA Explorer	25
Figure4.2	Dataset is Loaded	26
Figure4.3	Display First attribute of Dataset	27
Figure4.4	Shows of 2 nd attribute of dataset	27
Figure4.5	Description of 3 rd attribute of dataset	28
Figure4.6	Description of 4 th attribute frequency of dataset	28
Figure4.7	Shows 6 th attribute GPS Signals of dataset	29
Figure4.8	Description of 7 th attribute Satellite of dataset	30
Figure4.9	Description of 10 th attribute ACK Failure Count of dataset	30
Figure4.10	Graphical representation of all attributes	31
Figure4.11	IDTDMA algorithm output in NETBEANS	32
Figure4.12	Results of Proposed DT algorithm (IDTDMA)	33
Figure4.13	Results of Proposed DT algorithm (IDTDMA) to improve Correctly Classified Instance	34
Figure4.14	Results of Traditional DT algorithm (C45) to improve Correctly Classified Instance	35

