A Project Report
on

PERFORMANCE ANALYSIS OF CLASSIFICATION ALGORITHMS USING DIFFERENT DATASETS

Submitted for partial fulfilment of the requirements for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

By

Mr. G. Srihari (2451-15-733-149)

Under the guidance of

Mr. T. Srikanth

Assistant Professor

Department of CSE

M.V.S.R.E.C., Hyderabad.

Department of Computer Science and Engineering

M.V.S.R. ENGINEERING COLLEGE

(Affiliated to Osmania University & Recognized by AICTE)

Nadergul, Saroor Nagar Mandal, Hyderabad – 501 510

2018-19.

Department of Computer Science and Engineering

M.V.S.R. ENGINEERING COLLEGE

(Affiliated to Osmania University & Recognized by AICTE)

Nadergul, Saroor Nagar Mandal, Hyderabad – 501 510

CERTIFICATE

This is to certify that the project work entitled **"Performance Analysis of Classification Algorithms using Different Datasets"** is a bonafide work carried out by **Mr. G. Srihari** (**2451-15-733-149)** in partial fulfillment of academic requirements for the award of the degree of **BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING** from MVSR Engineering College, affiliated to OSMANIA UNIVERSITY, Hyderabad, under our guidance and supervision.

The results embodied in the report have not been submitted to any other university or institute for the award of any degree or diploma.

Internal Guide	**Head of the Department**
Mr. T. Srikanth	**Dr. Akhil Khare**
Assistant Professor	**Professor and Head**
Department of CSE	**Department of CSE**
MVSREC, Hyderabad.	**MVSREC, Hyderabad**.

DECLARATION

This is to certify that the work reported in the present project entitled "**Performance Analysis of Classification Algorithms using Different Datasets**" is a bonafide work done by me in the Department of Computer Science and Engineering, M.V.S.R. Engineering College, Osmania University. The report is based on the project work done entirely by me and not copied from any other source.

The results embedded in this project report have not been submitted to any other University or Institute for the award of any degree or diploma to the best of my knowledge and belief.

Mr. G. Srihari
(2451-15-733-149)

ACKNOWLEDGMENT

I take this opportunity to express my profound and sincere gratitude to all those who helped me in carrying out this project successfully.

I would like to thank my Internal Project Guide Mr. **T. Srikanth** , Assistant Professor, Department of Computer Science and Engineering, MVSR Engineering College for his useful suggestions, guidance and encouragement.

At the very outset, I am thankful to my principal **Dr. G. Kanaka Durga** and **Dr. Akhil Khare**, Professor and Head, Department of Computer Science and Engineering, MVSR Engineering College, Hyderabad for their consent to do the project work as a part of B.E Degree (CSE). I thank them for their valuable suggestions and advice throughout my stay at the college, during my project work.

I would like to thank my Project Co-ordinators Mrs. **D. Sirisha**, Assist Professor **Dr**. **M. Hanmandlu** Assoc. Proff ,Mr. **Md. Abdul Azeem** Assoc Proff and **Dr**. **Sesham Anand** Assoc proff for helping me complete project and for useful suggestions, guidance and encouragement.

I thank the teaching and non-teaching staff of CSE for extending their support.

Finally, I am thankful to my parents for their cooperation and support throughout all endeavors in my life. I sincerely acknowledge and thank all those who gave directly or indirectly their support in completion of this work

Mr. G. Srihari (2451-15-733-149)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VISION

- To impart technical education of the highest standards, producing competent and confident engineers with an ability to use computer science knowledge to solve societal problems.

MISSION

- To make learning process exciting, stimulating and interesting.
- To impart adequate fundamental knowledge and soft skills to students.
- To expose students to advanced computer technologies in order to excel in engineering practices by bringing out the creativity in students.
- To develop economically feasible and socially acceptable software.

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

The Program Educational Objectives of undergraduate program in Computer Science & Engineering are to prepare graduates who will:

1. Obtain strong fundamentals concepts, technical competency and problem solving skills to generate innovative solutions to engineering problems.
2. Continuously enhance their skills through training, independent inquiry, professional practices and pursue higher education or research by adapting to rapidly changing technology.
3. Advance in their professional careers including increased technical, multidisciplinary approach and managerial responsibility as well as attainment of leadership positions thus making them competent professionals at global level.
4. Exhibit commitment to ethical practices, societal contributions and lifelong learning.

(A) PROGRAM OUTCOMES(POs)

At the end of the program the students (Engineering Graduates) will be able to:

1. **Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis:** Identify, formulate, review research literature, and analyse complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/development of solutions:** Design solutions for complex engineering problems

and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. **Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.
6. **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentat ion, make effective presentations, and give and receive clear instructions.
11. **Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principle and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Lifelong learning:** Recognize the need for, and have the preparation and ability to engage in independent and life- long learning in the broadest context of technological change.

(B)PROGRAM SPECIFIC OUTCOMES (PSOs)

13. **Efficient coding:** an ability to analyse a problem, design the algorithm and optimally code itssolution.
14. **Software deployment:** an ability to identify & define computing requirements to test, implement and maintain a software product.

COURSE NAME: PROJECT COURSE
CODE: CS 414

Course Objectives:

1. Learn to survey the necessary domains for problemidentification
2. Learn the process of planning the complete lifecycle of aproject
3. Understand how to map requirements from a user into software specification.
4. Learn to apply concepts of software engineering for design of the identified real world problem
5. Improve the coding capabilities by implementing the various modules ofproject
6. Comprehend the suitable documentation procedure for a technical project.

Course Outcomes:

Code No.	Student will be able to
CS414.1	Summarize the survey of the recent advancements to infer the problem statement with applications towards society.
CS414.2	Design a software based solution within the scope of project.
CS414.3	Implement using contemporary technologies and tools.
CS414.4	Test and deploy the applications on real world environments.
CS414.5	Demonstrate qualities necessary for working in a team.
CS414.6	Generate a suitable technical document for the project.

ABSTRACT

Data Mining is a knowledge field that intersects domains from computer science and statistics, attempting to discover knowledge from databases in order to facilitate the decision making process. Classification is a Data Mining task that learns from a collection of cases in order to accurately predict the target class for new cases. Several machine learning techniques can be used to perform classification. In this report, I will be discussing about four different types of classification techniques(Decision Tree Classification, K nearest Neighbor Classification, Naive Bayes Classification, Random Forest Classification) which will be applied to data sets . I developed a GUI to analyze any dataset given to it by applying above techniques and builds a model for prediction for that particular dataset.

TABLE OF CONTENTS

PAGE NOS

Certificate ……. ii
Declaration…… iii
Acknowledgement…… iv
Vision and mission…… v
Abstract ……viii
Table of contents …….. ix
List of Figures…… xi
List of Tables ……xii

CHAPTER I
INTRODUCTION 01-04
1.1 Problem statement 03
1.2 Scope 03
1.3 Objectives 04
1.4 Proposed system 04

CHAPTER II
LITERATURE REVIEW 05-08
2.1 Introduction to python 05
2.2 Introduction to numpy, sklearn and pandas 06
2.3 Introduction to Tkinter 07
2.4 Research done before 07

CHAPTER III
METHODOLOGIES 09-16
3.1 KNN 09
3.2 Decision Tree (CART) 10
3.3 Naive Bayes 11
3.4 Random Forest 11
3.5 Datasets Used 12-15
3.5.1 Diabetes dataset 12
3.5.2 Car dataset 13
3.5.3 German Credit dataset 15
3.6 Accuracy Measures 16

CHAPTER IV
SYSTEM DESIGN AND IMPLEMENTATION 17-21
4.1 Data flow diagram 18
4.2 Class diagram 19
4.3 Package diagram 20
4.4 Hardware and Software requirements 21
4.5 Implementation 21

CHAPTER V
RESULT 22-24

CHAPTER VI
CONCLUSION AND FUTURE WORK 25

CHAPTER VII
REFERENCES 26

APPENDIX 28

LIST OF FIGURES

PAGE NOS

Figure	1.1	Data mining process	2
Figure	4.1	System Architecture	17
Figure	4.2	Dataflow Diagram	18
Figure	4.3	Class Diagram	19
Figure	4.4	Package Diagram	20
Figure	5.1	File Uploading	22
Figure	5.2	Performance Analysis	23
Figure	5.3	Comparing Accuracy Graphs	23
Figure	5.4	Prediction Part	24
Figure	5.5	Result of Prediction	24

LIST OF TABLES

PAGE NOS

Table	3.1	Diabetes dataset Descriptor	13
Table	3.2	Diabetes dataset Attributes	13
Table	3.3	Car dataset Model Evaluation	14
Table	3.4	Car dataset Descriptor	14
Table	3.5	Car dataset Attributes	14
Table	3.6	German Credit data Descriptor	15
Table	3.7	German Credit data Attributes	15