Criterion 1: Curricular Aspects

1.3 Curriculum Enrichment

1.3.4.1: Number of students undertaking field projects / internships / student projects

Programme Name: M.E Computer Science and Engineering.

Sl.No.	Description	Page Number
1	Student Projects	1 - 76

Criterion 1: Curricular Aspects

1.3 Curriculum Enrichment

1.3.4.1: Number of students undertaking field projects / internships / student projects

Programme Name: M.E Computer Science and Engineering

Student Projects Proof

A DYNAMIC APPROACHES TOWARDS SENSITIVE LABEL PRIVACY PRESERVATION WITH AUTOMIZATION FOR DATA PUBLISHING

A PROJECT REPORT PHASE I

Submitted by

KAVIBHARANI S (927621MCS003)

in partial fulfillment for the award of the degree

of

MASTER OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

M.KUMARASAMY COLLEGE OF ENGINEERING
(An Autonomous Institution, Affiliated to Anna University, Chennai)

KARUR -639 113

DECEMBER 2022

M.KUMARASAMY COLLEGE OF ENGINEERING

(An Autonomous Institution, Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this project report **"A DYNAMIC APPROACHES TOWARDS SENSITIVE LABEL PRIVACY PRESERVATION WITH ANATOMIZATION FOR DATA PUBLISHING"** is the bonafide work of **"KAVIBHARANI S (927621MCS2003)"** who carried out the project work under my supervision.

Signature

Dr.S.THILAGAMANI, M.E., Ph.D.,

HEAD OF THE DEPARTMENT,

Department of Computer Science and Engineering,
M.Kumarasamy College of Engineering,
Thalavapalayam, Karur – 639 113.

Signature

Dr.S.SUJANTHI, M.E.,Ph.D.,

SUPERVISOR,

Department of Computer Science and Engineering,
M.Kumarasamy College of Engineering,
Thalavapalayam, Karur – 639 113.

Submitted for Project Work Phase I viva voce examination held on 28.12.2022

INTERNAL EXAMINER

28/12/2022

EXTERNAL EXAMINER

DECLARATION

I affirm that the Project report titled "ENHANCING **PACKET DELIVERY RATIO USING GRAY HOLE ATTACK IN MANET**" being submitted in partial fulfillment for the award of **Master of Engineering in Computer Science and Engineering**, is the original work carried out by me. It has not formed the part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

S. Kavibharani

KAVIBHARANI S (927621MCS2003)

I certify that the declaration made by the above candidate is true to the best of my knowledge.

Name & Signature of the supervisor with date

Dr. S. SUJANTHI

ACKNOWLEDGEMENT

Behind every achievement lies an unfathomable sea of gratitude to those who actuated it, without them it would have never came into existence, to them we lay the word of gratitude imprinted within us.

I would like to express my profuse gratitude to **Thiru.M.Kumarasamy, Chairman** and **Dr.K.Ramakrishnan, Secretary** of our college for providing extra ordinary infrastructure, which helped me in the completion of the project Phase I in time.

I wish to express my sincere thanks to our respected **Dr.N.Ramesh Babu, M.E., Ph.D., Principal,** for all the blessing and help provided during the period of project work.

I explore my gratitude to **Dr.S.Thilagamani, M.E., Ph.D., Head of the Department, Computer Science and Engineering**, for permitting me to undertake this project as part of the curriculum.

I am indebted to my project supervisor **Dr.B.Padmini Devi, M.E., Ph.D., Associate Professor, Computer Science and Engineering,** for her constant help and creative ideas over the period of project work.

I am thankful to my faculty members of CSE department for their continuous direction and guidance and timely support extended for me to carry out my project successfully.

S. Kavibharani

DATE: 28.12.2022 [KAVIBHARANI S]

AN ACCURATE PREDICTION OF BITCOIN PRICE USING DATA SCIENCE AND DEEP LEARNING

A PROJECT REPORT PHASE I

Submitted by

NAGA KANNIKA S (927621MCS004)

in partial fulfillment for the award of the degree

of

MASTER OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

M.KUMARASAMY COLLEGE OF ENGINEERING
(An Autonomous Institution, Affiliated to Anna University, Chennai)
KARUR – 639 113

DECEMBER 2022

M.KUMARASAMY COLLEGE OF ENGINEERING

(An Autonomous Institution, Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this project report **"AN ACCURATE PREDICTION OF BITCOIN PRICE USING DATA SCIENCE AND DEEP LEARNING"** is the bonafide work of **"NAGA KANNIKA S (927621MCS004)"** who carried out the project work under my supervision.

Signature

Dr.S.THILAGAMANI, M.E., Ph.D.,

HEAD OF THE DEPARTMENT,

Department of Computer Science and Engineering,
M.Kumarasamy College of Engineering,
Thalavapalayam, Karur – 639 113.

Signature

Dr.B.PADMINI DEVI, M.E.,Ph.D.,

SUPERVISOR,

Department of Computer Science and Engineering,
M.Kumarasamy College of Engineering,
Thalavapalayam, Karur – 639 113.

Submitted for Project Work Phase I viva voce examination held on 28.12.2022

INTERNAL EXAMINER

20/12/2022

EXTERNAL EXAMINER

DECLARATION

I affirm that the Project report titled "AN ACCURATE PREDICTION OF BITCOIN PRICE USING DATA SCIENCE AND DEEP LEARNING" being submitted in partial fulfillment for the award of **Master of Engineering in Computer Science and Engineering**, is the original work carried out by me. It has not formed the part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

NAGA KANNIKA S (927621MCS004)

I certify that the declaration made by the above candidate is true to the best of my knowledge.

Name & Signature of the supervisor with date

DR. B. PADMINI DEVI

ACKNOWLEDGEMENT

Behind every achievement lies an unfathomable sea of gratitude to those who actuated it, without them it would have never came into existence, to them we lay the word of gratitude imprinted within us.

I would like to express my profuse gratitude to **Thiru.M.Kumarasamy, Chairman** and **Dr.K.Ramakrishnan, Secretary** of our college for providing extra ordinary infrastructure, which helped me in the completion of the project Phase I in time.

I wish to express my sincere thanks to our respected **Dr.N.Ramesh Babu, M.E., Ph.D., Principal,** for all the blessing and help provided during the period of project work.

I explore my gratitude to **Dr.S.Thilagamani, M.E., Ph.D., Head of the Department, Computer Science and Engineering**, for permitting me to undertake this project as part of the curriculum.

I am indebted to my project supervisor **Dr.B.Padmini Devi, M.E., Ph.D., Associate Professor, Computer Science and Engineering,** for her constant help and creative ideas over the period of project work.

I am thankful to my faculty members of CSE department for their continuous direction and guidance and timely support extended for me to carry out my project successfully.

DATE: 28.12.2022

28/12/2022

[NAGA KANNIKA S]

ABSTRACT

Bitcoin uses a peer-to-peer technology to operate with no central authority or banks. Bitcoin is open-source; its design is public, nobody owns or controls Bitcoin and everyone can take part. Digital currency brings into use as open source software in pseudonymous creator Satoshi Nakamoto It is a crypto currency, so-called because it uses cryptography to control the creation and transfer of money. The goal of this work is to compare the accuracy of bitcoin price in USD prediction based on Long Short-term Memory (LSTM) network with self-attention. Real-time price data is collected by Pycurl from Bitfine. LSTM model is implemented by Keras and TensorFlow. The proposed model used in this work is mainly to present a classical comparison of time series forecasting, as expected, it could make efficient prediction limited in short-time interval, and the outcome depends on the time period. The LSTM could reach a better performance, with extra, indispensable time for model training, especially via CPU.

Programme: Master of Engineering- Computer Science and Engineering

Vision of the Department

To achieve education and research excellence in Computer Science and Engineering.

Mission of the Department

M1: To excel in academic through effective teaching learning techniques

M2: To promote research in the area of computer science and engineering with the focus on innovation

M3: To transform students into technically competent professionals with societal and ethical responsibilities

Program Outcomes (POS)

PO1: An ability to independently carry out research / investigation and development work to solve practical problems.

PO2: An ability to write and present a substancial technical report/document.

PO3: Students should be able to demonstrate a degree of mastery over the area asper the specialization of the program .The mastery should be at a level higher than than the requirements in the appropriate bachelor program.

PO4: Ability to discriminate,evaluate,analyze and synthesize existing and new knowledge and integration of the same for enhancement of knowledge in Computer Science and Engineering.

PO5: Ability to think laterally and originally to identify,formulate and solve an engineering problem in Computer Science and Engineering and effectively utilize appropriate scientific and engineering techniques and methodologies in the problem solving process.

PO6: Ability to apply the tools from optimization, probability,statistics,simulation and engineering economic analysis,including fundamental application of the tools in IT industry invoving uncertainty and scarce or expensive resources.

Program Educational Objectives (PEOs)

PEO 1: To empower graduates to identify,create and solve computing problem by applying their knowledge of computing principles and mathematical theory to develop sustainable solutions to current and future computing problems.

PEO 2: To develop research attitude in graduates and to explore it for higher education Endeavors and constantly upgrade their skills with an attitude towards lifelong learning.

PEO 3: To facilitate graduates to acquire skills to communicate effectively with the society and contribute to the betterment of the society as a commited technical personnel.

PROJECT MAPPED WITH PO AND PEO

ABSTRACT	PO's MAPPED	PEO's MAPPED
Bitcoin uses a peer-to-peer technology to operate with no central authority or banks. Bitcoin is open-source; its design is public, nobody owns or controls Bitcoin and everyone can take part. Digital currency brings into use as open source software in pseudonymous creator Satoshi Nakamoto It is a crypto currency, so-called because it uses cryptography to control the creation and transfer of money. The goal of this work is to compare the accuracy of bitcoin price in USD prediction based on Long Short-term Memory (LSTM) network with self-attention. Real-time price data is collected by Pycurl from Bitfine. LSTM model is implemented by Keras and TensorFlow. The proposed model used in this work is mainly to present a classical comparison of time series forecasting, as expected, it could make efficient prediction limited in short-time interval, and the outcome depends on the time period. The LSTM could reach a better performance, with extra, indispensable time for model training, especially via CPU.	PO1(L) PO2(M) PO3(H) PO4(M) PO5(M) PO6(M)	PEO1(M) PEO(H) PEO(M)

SUPERVISOR

HEAD OF DEPARTMENT

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	i
	TABLE OF CONTENTS	ii
	LIST OF TABLES	v
	LIST OF FIGURES	Vi
	LIST OF ABBREVIATIONS	vii
1	INTRODUCTION	1
	1.1 TRANSACTIONS	4
	1.2 OWNERSHIP	5
	1.3 MINING	6
	1.4 DECENTRALIZATION	7
	1.5 PRIVACY AND FUNGIBILITY	8
	1.6 HISTORY	8
	1.7 DATA MINING FOR BITCOIN PREDICTION	14
	1.8 DATAPREPROCESSING	14
	1.9 DATA MININGTECHNIQUES	15
	1.10 DATA MININGTASKS	16
	1.11 DATA MININGAPPLICATIONS	18
	1.12 STATISTICAL ANDDATMININGTOOLS	18

CHAPTER NO	TITLE	PAGE NO
2	LITERATURE REVIEW	21
	2.1 PROBLEM STATEMENT	30
3	EXISTING SYSTEM	31
	3.1 CLASSIFICATION AND PREDICTION ISSUES	31
	3.2 COMPARISON OF CLASSIFICATION AND PREDICTION METHODS	32
	3.3 TYPES OF ALGORITHMS IN DATA MINING	32
	3.3.1 STATISTICAL PROCEDURE BASED APPROACH	33
	3.3.2 MACHINE LEARNING-BASED APPROACH	33
	3.3.3 NEURAL NETWORK	33
	3.3.4 CLASSIFICATION ALGORITHMS IN DATA MINING	34
	3.3.5 ID3 ALGORITHM	34
	3.3.6 C4.5 ALGORITHM	35
	3.3.7 K NEAREST NEIGHBORS ALGORITHM	35
	3.3.8 NAÏVE BAYES ALGORITHM	36
	3.3.9 SVM ALGORITHM	37
	3.3.10 ANN ALGORITHM	38
	3.3.11 SUPPORT VECTOR MACHINES	39

CHAPTER NO	TITLE	PAGE NO
4	**PROPOSED SYSTEM**	42
	4.1 DATA SET COLLECTION	42
	4.2 PRE-PROCESSING	42
	4.3 TRAINING	42
	4.4 PREDICTION	42
	4.5 CONSTRUCTION OF ATTENTION-LSTM MODEL	43
	4.5.1 LSTM MODEL	43
	4.6 PROPOSED ALGORITHM 1	47
	4.7 EVALUATION	48
	4.7.1 CONFUSION MATRIX	48
	4.8 AREA UNDER CURVE	49
5	**SOFTWARE SPECIFICATION**	51
	5.1 PYTHON	51
	5.1.1 WHY PYTHON?	51
	5.1.2 INSTALLATION	52
6	**CONCLUSION**	53
	REFERENCES	54

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
2.1	Comparison Table For Literature Review	26

SYMBOLS

□	Bitcoin
$	Dollar

LIST OF FIGURES

FIGURE NO.	FIGURE NAME	PAGE NO.
1.1	Data structure of blocks in the ledger	2
1.2	Number of bitcoin transactions per month, semi logarithmic plot	2
1.3	Number of unspent transaction outputs	3
1.4	Bitcoin ownership	5
1.5	Bitcoin price	12
1.6	Data mining as a step in the process of knowledge discovery	14
1.7	Data mining tasks	16
1.8	Statistical and data mining tools	19
3.1	Date mining algorithms	32
4.1	Recurrent neural network (RNN)	43
4.2	The structure of LSTM cell	43
4.3	The process of adding an attention mechanism to the LSTM model	45
4.4	The internal structure of attention model	45
4.5	Attention-LSTM network architecture	47
4.6	Confusion Matrix	48

LIST OF ABBREVIATIONS

POW	Proof Of Work
FINCEN	Financial Crimes Enforcement Network
MSB	Money Service Business
Seg Wit	Segregated Witness
CEC	Center For Cition Studies
KDD	Knowledge Discovery In Databases
SVM	Support Vector Machine
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Roof Mean Squared Error
LSTM	Long-short Term Memory
ARIMAX	Auto Regressive Integrated Moving Average with exogenous input
BTC	Bitcoin
MBTC	Millibitcoin
RNN	Recurrent Neural Network
BNN	Bayesian Neural Network
MAPE	Mean Absolute Percentage
KNN	K Nearest Nearest Neighbour
ANN	Artificial Neural Networks
AUC	Area Under Curve

QUALITY ANALYSIS OF EGGS USING IMAGE PROCESSING WITH INTERPRETABLE MACHINE LEARNING MODEL

A PROJECT REPORT PHASE I

Submitted by

Mrs.G.RAMYA SHRI (927621MCS006)

in partial fulfillment for the award of the degree

of

MASTER OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

M.KUMARASAMY COLLEGE OF ENGINEERING

(An Autonomous Institution, Affiliated to Anna University, Chennai)

KARUR – 639 113

DECEMBER 2022

AN ANALYSIS OF DIFFERENT DATA SAMPLING METHOD FOR MEDICAL DATA USING MACHINE LEARNING

A PROJECT REPORT PHASE I

Submitted by

SHANMUGASUNDARAM P (927621MCS007)

in partial fulfillment for the award of the degree

of

MASTER OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

M.KUMARASAMY COLLEGE OF ENGINEERING
(An Autonomous Institution, Affiliated to Anna University, Chennai)
KARUR – 639 113

DECEMBER 2022

M.KUMARASAMY COLLEGE OF ENGINEERING

(An Autonomous Institution, Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this project report "**AN ANALYSIS OF DIFFERENT DATA SAMPLING METHOD FOR MEDICAL DATA USING MACHINE LEARNING**" is the bonafide work of "SHANMUGASUNDARAM P (**927621MCS007**)" who carried out the project work under my supervision.

Signature

Dr.S.THILAGAMANI, M.E., Ph.D.,

HEAD OF THE DEPARTMENT,

Department of Computer Science and Engineering,
M.Kumarasamy College of Engineering,
Thalavapalayam, Karur – 639 113.

Signature

Mr.V. Mani, M.E,

SUPERVISOR,

Department of Computer Science and Engineering,
M.Kumarasamy College of Engineering,
Thalavapalayam, Karur – 639 113.

Submitted for Project Work Phase I viva voce examination held on 28/12/2022.

INTERNAL EXAMINER

28/12/2022

EXTERNAL EXAMINER

DECLARATION

I affirm that the Project report titled "**AN ANALYSIS OF DIFFERENT DATA SAMPLING METHOD FOR MEDICAL DATA USING MACHINE LEARNING**" being submitted in partial fulfillment for the award of **Master of Engineering in Computer Science and Engineering**, is the original work carried out by me. It has not formed the part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

28/12/22.

SHANMUGASUNDARAM P (927621MCS007)

I certify that the declaration made by the above candidate is true to the best of my knowledge.

Name & Signature of the supervisor with date

Mr. V. Mani

ACKNOWLEDGEMENT

Behind every achievement lies an unfathomable sea of gratitude to those who actuated it, without them it would have never came into existence, to them we lay the word of gratitude imprinted within us.

I would like to express my profuse gratitude to **Thiru.M.Kumarasamy, Chairman** and **Dr.K.Ramakrishnan, Secretary** of our college for providing extra ordinary infrastructure, which helped me in the completion of the project Phase I in time.

I wish to express my sincere thanks to our respected **Dr.N.Ramesh Babu, M.E., Ph.D., Principal,** for all the blessing and help provided during the period of project work.

I explore my gratitude to **Dr.S.Thilagamani, M.E., Ph.D., Head of the Department, Computer Science and Engineering**, for permitting me to undertake this project as part of the curriculum.

I am indebted to my project supervisor **Mr.V.Mani, M.E., Associate Professor, Computer Science and Engineering,** for his constant help and creative ideas over the period of project work.

I am thankful to my faculty members of CSE department for their continuous direction and guidance and timely support extended for me to carry out my project successfully.

DATE: 08/12/2022 [SHANMUGASUNDARAM P]

Program: Master of Engineering- Computer Science and Engineering

Vision of the Department

To achieve education and research excellence in Computer Science and Engineering

Mission of the Department

M1: To excel in academic through effective teaching learning techniques

M2: To promote research in the area of computer science and engineering with the focus on innovation

M3: To transform students into technically competent professionals with societal and ethical responsibilities

Program Outcomes (POS)

PO1: An ability to independently carry out research / investigation and development work to solve practical problems.

PO2: An ability to write and present a substancial technical report/document.

PO3: Students should be able to demonstrate a degree of mastery over the area asper the specialization of the program .The mastery should be at a level higher than than the requirements in the appropriate bachelor program.

PO4: Ability to discriminate,evaluate,analyze and synthesize existing and new knowledge and integration of the same for enhancement of knowledge in Computer Science and Engineering.

PO5: Ability to think laterally and originally to identify,formulate and solve an engineering problem in Computer Science and Engineering and effectively utilize appropriate scientific and engineering techniques and methodologies in the problem solving process.

PO6: Ability to apply the tools from optimization, probability,statistics,simulation and engineering economic analysis,including fundamental application of the tools in IT industry invoving uncertainty and scarce or expensive resources

Program Educational Objectives (PEOs)

PEO 1: To empower graduates to identify,create and solve computing problem by applyingtheir knowledge of computing principles and mathematical theory to develop sustainable solutions to current and future computing problems.

PEO 2: To develop research attitude in graduates and to explore it for higher educationEndeavors and constantly upgrade their skills with an attitude towards lifelonglearning.

PEO 3: To facilitate graduates to acquire skills to communicate effectively with the societyand contribute to the betterment of the society as a committed technical personnel.

PROJECT MAPPED WITH PO AND PEO

ABSTRACT	PO's MAPPED	PEO's MAPPED
ıy's era of internet, the amount of data generation is growing on sing. With the development of artificial intelligence, big data ication technology provides the advantageous help for the medicine ry diagnosis research. While due to the different conditions in the nt sample collection, the medical big data is often imbalanced. nced data typically refers to a condition in which several data s in a certain problem is not equally distributed, thereby leading to lerrepresentation of one or more classes in the dataset. onal classification algorithms usually assume that the number of s in each class is similar and their misclassification cost during g is equal. However, the misclassification cost of patient samples is than that of healthy person samples. Therefore, how to increase the ication of patients without affecting the classification of healthy uals is an urgent problem. aightforward method to solve this problem is the resampling method ing records to the minority class or deleting ones from the majority In this paper, we have experimented with different SMOTE mpling and undersampling methods. rpose of this work is to balance the imbalanced data using different ng techniques. We use two medical dataset i.e **PIMA INDIAN ETICS DATASET** and **CHRONIC KIDNEY DISEASE SET** download from kaggle repository. in this project, split in to two phases. One is data sampling and other Prediction model. project we used different data sampling methods like **SMOTE, K-SMOTE, SMOTENC (SMOTE for Nominal and Continuous) MOTEN (SMOTE for Nominal).** getting modified data sampling dataset, to apply the different ne learning algorithms i.e **Decision tree, Random Forest, SVM and** to predict the prediction of Chronic Kidney Disease and diabetic e in early stage. on **accuracy, precision and Recall** value from implemented tested ne learning model to find out the best Sampling as well as machine ıg algorithms.	PO1(L) PO2(M) PO3(H) PO4(M) PO5(M) PO6(M)	PEO1(M) PEO(H) PEO(M)

ABSTRACT

In today's era of internet, the amount of data generation is growing on increasing. With the development of artificial intelligence, big data classification technology provides the advantageous help for the medicine auxiliary diagnosis research. While due to the different conditions in the different sample collection, the medical big data is often imbalanced.

Imbalanced data typically refers to a condition in which several data samples in a certain problem is not equally distributed, thereby leading to the underrepresentation of one or more classes in the dataset.

Traditional classification algorithms usually assume that the number of samples in each class is similar and their misclassification cost during training is equal. However, the misclassification cost of patient samples is higher than that of healthy person samples. Therefore, how to increase the identification of patients without affecting the classification of healthy individuals is an urgent problem.

The straightforward method to solve this problem is the resampling method by adding records to the minority class or deleting ones from the majority class. In this paper, we have experimented with different SMOTE oversampling and undersampling methods.

The purpose of this work is to balance the imbalanced data using different sampling techniques. We use two medical dataset i.e **PIMA INDIAN DIABETICS DATASET** and **CHRONIC KIDNEY DISEASE DATASET** download from kaggle repository.

So that in this project, split in to two phases. One is data sampling and other one is Prediction model.

In this project we used different data sampling methods like **SMOTE, K-means SMOTE, SMOTENC (SMOTE for Nominal and Continuous) and SMOTEN (SMOTE for Nominal).**

After getting modified data sampling dataset, to apply the different Machine learning algorithms i.e **Decision tree, Random Forest, SVM and KNN** to predict the prediction of Chronic Kidney Disease and diabetic disease in early stage.

Based on **accuracy, precision and Recall** value from implemented tested machine learning model to find out the best Sampling as well as machine learning algorithms.

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	**i**
	TABLE OF CONTENTS	**ii**
	LIST OF TABLES	**iii**
	LIST OF FIGURES	**iv**
	LIST OF ABBREVIATIONS	**v**
1	**INTRODUCTION**	**1**
2	**LITERATURE REVIEW**	**7**
3	**EXISTING SYSTEM**	**26**
4	**PROBLEM DESCRIPTION**	**27**
5	**PROPOSED SYSTEM**	**28**
	5.1 INTRODUCTION	**28**
	5.2 METHODOLOGY	**28**
	5.2.1 DATASET DECRIPTION	**28**
	5.2.2 DATA ARECHIECTURE	**29**
	5.2.3 SAMPLING TECHNIQUES DETAILS	**20**
	5.3 MODULES	**32**
	5.3.1 MODULES DETAILS	**32**
	5.3.4 PERFORMANCES EVALUATION MEASURE	**33**
6	**CONCLUSION**	**35**
	REFERENCES	**36**

LIST OF TABLES

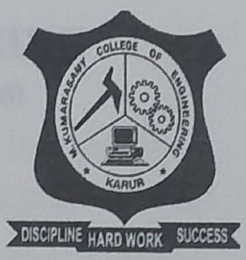
TABLE NO.	TITLE	PAGE NO.
2.1	Comparison Table For Literature Review	28

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
1.1	Different between underdamping and oversampling	2
1.2	CKD disease description	3
5.1	Data architecture diagram	29
5.2	Working principal of SMOTE	30
5.3	Working principal of ADAYSN	31
5.4	Working principal of SMOTE +TOMEK	32

LIST OF ABBREVIATIONS

ML	Machine Learning
CKD	Chronic Kidney Disease
SMOTE	Synthetic Minority Oversampling Technique
ADASYN	Adaptive Synthetic Sampling Approach
ACK	Acknowledgement



A DYNAMIC APPROCHES TOWARDS SENSITIVE LABEL PRIVACY PRESERVATION WITH ANATMIZTION FOR DATA PUBLISHING

A PROJECT REPORT PHASE II

Submitted by

S.KAVIBHARANI (927621MCS003)

in partial fulfillment for the award of the degree

of

MASTER OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

M.KUMARASAMY COLLEGE OF ENGINEERING

(An Autonomous Institution, Affiliated to Anna University, Chennai)

KARUR – 639 113

ANNA UNIVERSITY :: CHENNAI 600 025

APRIL 2023

M.KUMARASAMYCOLLEGEOF ENGINEERING

(AnAutonomous Institution, Affiliated toAnna University,Chennai)

BONAFIDE CERTIFICATE

Certified that this project report **"A DYNAMIC APPROCHES TOWARDS SENSTIVE LABEL PRIVACY PRESERVATION WITH AUTOMIZATION FORDATA PUBLISHING"** is the bonafide work of **S.KAVIBHARANI (927621MCS003)"** who carried out the project work during the academic year 2022-2023 under my supervision. Certified furtherthat other best of my knowledge the work reported herein does not form partof any other thesis or dissertation on the basis of which a degree or award was occasion on this or any other candidate.

Signature
Dr.S.SUJANTHI, M.E.,Ph.D.,
SUPERVISOR,
DepartmentofComputerScienceand Engineering,
M.KumarasamyCollegeOfEngineering,
Thalavapalayam,Karur–639113.

Signature
Dr.M.MURUGASEN,M.E., Ph.D.,
HEADOFTHE DEPARTMENT,
DepartmentofComputerScience and Engineering,
M.KumarasamyCollege OfEngineering,
Thalavapalayam,Karur –639113.

SubmittedforProject WorkPhaseIIvivavoceexamination held on 12.4.2023

INTERNALEXAMINER

EXTERNALEXAMINER

DECLARATION

I affirm that the Project report titled **" A DYNAMIC APPROCHES TOWARDS SENSITIVE LABEL PRIVACY PRESERVATION WITH ANATMIZTION FOR DATA PUBLISHING"** being submitted in partial fulfillment for the award of Master of Engineering in Computer Science and Engineering, is the original work carried out by me. It has not formed the part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

S.Kavibharani

S.KAVIBHARANI (927621MCS003)

I certify that the declaration made by the above candidate is true to the best of my knowledge.

14/4/23

Name & Signature of the supervisor with date

[Dr. S. SUBANTHI]

ACKNOWLEDGEMENT

Behind every achievement lies an unfathomable sea of gratitude to those who actuated it, without them it would have never came into existence, to them we lay the word of gratitude imprinted within us.

I would like to express my profuse gratitude to **Thiru.M.Kumarasamy, Chairman** and **Dr.K.Ramakrishnan, Secretary** of our college for providing extra ordinary infrastructure, which helped me in the completion of the project Phase II in time.

I wish to express my sincere thanks to our respected **Dr.B.S.Murugan, M.E., Ph.D., Principal,** for all the blessing and help provided during the period of project work.

I explore my gratitude to **Dr.M.Murugasen, M.E., Ph.D., Head of the Department, Computer Science and Engineering** for permitting me to undertake this project as part of the curriculum.

I am indebted to my project supervisor **Dr.S.Sujanthi,M.E.,Ph.D., Assistant Professor, Computer Science and Engineering,** for her constant help and creative ideas over the period of project work .

I am thankful to **my faculty members of CSE department** for their continuous direction and guidance and timely support extended for me to carry out my project successfully.

Words are boundless to thank **Our Parents and Friends** for their constant encouragement to complete this project successfully.

S. Kavibharani

DATE: 12.04.2023 [S.KAVIBHARANI]

M.KUMARASAMY COLLEGE OF ENGINEERING

Programme: Master of Engineering - Computer Science and Engineering

Vision of the Department

To achieve education and research excellence in Computer Science and Engineering

Mission of the Department

M1: To excel in academic through effective teaching learning techniques

M2: To promote research in the area of computer science and engineering with the focus on innovation

M3: To transform students into technically competent professionals with societal and ethical responsibilities

Program Outcomes (POS)

PO1: An ability to independently carry out research / investigation and development work to solve practical problems.

PO2: An ability to write and present a substancial technical report/document.

PO3: Students should be able to demonstrate a degree of mastery over the area asper the specialization of the program .The mastery should be at a level higher than than the requirements in the appropriate bachelor program.

PO4: Ability to discriminate,evaluate,analyze and synthesize existing and new knowledge and integration of the same for enhancement of knowledge in Computer Science and Engineering.

PO5: Ability to think laterally and originally to identify,formulate and solve an engineering problem in Computer Science and Engineering and effectively utilize appropriate scientific and engineering techniques and methodologies in the problem solving process.

PO6: Ability to apply the tools from optimization, probability,statistics,simulation and engineering economic analysis,including fundamental application of the tools in IT industry invoving uncertainty and scarce or expensive resources.

Program Educational Objectives (PEOs)

PEO 1: To empower graduates to identify,create and solve computing problem by applying their knowledge of computing principles and mathematical theory to develop sustainable solutions to current and future computing problems.

PEO 2: To develop research attitude in graduates and to explore it for higher education Endeavors and constantly upgrade their skills with an attitude towards lifelong learning.

PEO 3: To facilitate graduates to acquire skills to communicate effectively with the society and contribute to the betterment of the society as a commited technical personnel.

ABSTRACT

Data in its original form, however, typically contain sensitive information about the individuals. Directly publishing raw data will violate the privacy of people involved. Consequently, it becomes increasingly important to preserve the privacy of published data. An attacker is apt to identify an individual from the published tables, with attacks through the record linkage, attribute linkage, table linkage or probabilistic attack. Although algorithms based on generalization and suppression has been proposed to protect the sensitive attributes and resist these multiple types of attacks, they often suffer from large information loss by replacing specific values with more general ones. Alternatively, anatomization and permutation operations can de-link the relation between attributes without modifying them. In this paper, to propose a scheme Sensitive Label Privacy Preservation with Anatomization (SLPPA) was to protect the privacy of published data. The security analysis shows our scheme is provably secure, and the performance evaluation demonstrates the overhead of data dynamics and the dispute arbitrations are reasonable .

PROJECT MAPPED WITH PO AND PEO

ABSTRACT	PO'sMAPPED	PEO's MAPPED
Data in its original form, hotover, typically contain sensitive information about individuals. Directly publishing raw data will violate the privacy of people involved. Consequently, it becomes increasingly important to preserve the privacy of published data. An attacker is apt to identify an individual from the published tables, with attacks through the record linkage, attribute linkage, table linkage or probabilistic attack. Although algorithms based on generalization and suppression have been proposed to protect the sensitive attributes and resist these multiple types of attacks, they often suffer from large information loss by replacing specific values with more general ones. Alternatively, anatomization and permutation operations can de-link the relation bettoen attributes without modifying them. In this paper, To propose a scheme Sensitive Label Privacy Preservation with Anatomization (SLPPA) to protect the privacy of published data. The security analysis shows our scheme is provably secure, and the performance evaluation demonstrates the overhead of data dynamics and dispute arbitration are reasonable.	PO1(L) PO2(M) PO3(H) PO4(M) PO5(M) PO6(M)	PEO1(M) PEO2(H) PEO3(M)

L-Low M-Medium H-High

SUPERVISOR

HEAD OF THE DEPARTMENT

TABLE OF CONTENT

CHAPTER NO.	TITLE	PAGE NO
	ABSTRACT	i
	PROJECT MAPPED WITH PO AND POE	iiii
	TABLE OF CONTENTS	iii
	LIST OF TABLES	v
	LIST OF FIGURES	vi
	LIST OF ABBREVIATIONS	vii
1.	INTRODUCTION	1
	1.1 Cloud computing	1
	1.2 Cloud storage	2
	1.3 Cloud data	3
2.	LITERATURE REVIEW	4
3.	EXSISTING SYSTEM	17
	3.1 introduction	17
	3.2 Disadvantage	17
4.	PROPOSED SYSTEM	18
	4.1 Introduction	18
	4.2Advantages of proposed system	20
5.	SYSTEM REQUIREMENTS	23
6.	SYSTEM IMPLEMENTATION	24
	6.1 Modules	24
	6.2 Client Module	24
	6.3 Data upload and validation	24
7.	RESULT & DISCUSSION	27
8.	CONCLUSION	28

APPENDIX 1: SCREENSHOT 29

APPENDIX 2: SOURCE CODE 31

REFERENCES 54

LIST OF PUBLICATION 56

LISTOFTABLES

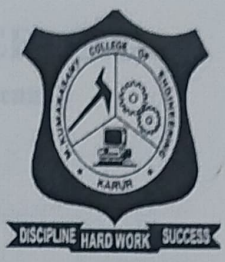
TABLE NO.	TABLE NAME	PAGE NO.
3.1	COMPARISONTABLEOFREFERENCEPAPER	13

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
4.1	Architecture of proposed system	19
6.1	Suppression Table	26
6.2	Group division graph	26
7.1	Result Analyses of Adult and Census-Income	27
A.1	New User Register	29
A.2	User Login	29
A.3	Upload File	30
A.4	DATASET	30

LIST OF ABBREVIATIONS

SLPPA	Service Location Protocol Profit Authentication
SLP	Stop Loss Payment
QOS	Quality Of Service
RREQ	Route Request
RREP	Route Reply
RERR	Route Error
SAS	Software As A Service
AODV	Ad-Hoc-On-Demand Distance Vector
PAS	Platform As A Service
IDAD	Intrusion Detection Using Anomaly Detection
DTN	Delay Tolerant Network
ACK	Acknowledgement
VANET	Vehicular Ad Hoc Network
SPAN	Smart Phone Ad Hoc Networks
IMANET	Internet Based Mobile Ad Hoc Network
IOT	Internet Of Things
IDS	Intrusion Detection System



AN ACCURATE PREDICTION OF BITCOIN PRICE USING DATA SCIENCE AND DEEP LEARNING

PROJECT REPORT

Submitted by

Mrs.S.NAGA KANNIKA (927621MCS004)

in partial fulfillment for the award of the degree

of

MASTER OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

M.KUMARASAMY COLLEGE OF ENGINEERING

KARUR – 639 113

ANNA UNIVERSITY : CHENNAI 600025

APRIL 2023

M.KUMARASAMY COLLEGE OF ENGINEERING

(An Autonomous Institution, Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this project report **"AN ACCURATE PREDICTION OF BITCOIN PRICE USING DATA SCIENCE AND DEEP LEARNING"** is the bonafide work of "**NAGA KANNIKA S (927621MCS004)**" who carried out the project work under my supervision.Certified futher that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an early occasion on this or any other candidate.

Signature

Dr.B.PADMINI DEVI, M.E., Ph.D.,

SUPERVISOR,

Department of Computer Science and Engineering,

M.Kumarasamy College of Engineering,

Thalavapalayam, Karur – 639 113.

Signature

Dr.M.MURUGESAN M.E., Ph.D.,

HEAD OF THE DEPARTMENT,

Department of Computer Science and Engineering,

M.Kumarasamy College of Engineering,

Thalavapalayam, Karur – 639 113.

Submitted for Project Work viva voce examination held on 12.04.2023

INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION

I affirm that the Project report titled **"AN ACCURATE PRDICTION OF BITCOIN PRICE USING DATA SCIENCE AND DEEP LEARNING"** being submitted in partial fulfillment for the award of Master of Engineering in Computer Science and Engineering, is the original work carried out by me. It has not formed the part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

S.NAGA KANNIKA (927621MCS004)

I certify that the declaration made by the above candidate is true to the best of my knowledge.

12/4/25

Name & Signature of the supervisor with date

Dr. B. Padmini Devi

ACKNOWLEDGEMENT

Behind every achievement lies an unfathomable sea of gratitude to those who actuated it, without them it would have never came into existence, to them we lay the word of gratitude imprinted within us.

I would like to express my profuse gratitude to **Thiru.M.Kumarasamy, Chairman** and **Dr.K.Ramakrishnan, Secretary** of our college for providing extra ordinary infrastructure, which helped me in the completion of the project Phase I in time.

I wish to express my sincere thanks to our respected **Dr.N.Ramesh Babu, M.E., Ph.D., Principal,** for all the blessing and help provided during the period of project work.

I explore my gratitude **to Dr.M.Murugesan M.E., Ph.D., Head of the Department, Computer Science and Engineering,** for permitting me to undertake this project as part of the curriculum.

I am indebted to my project supervisor **Dr.B.Padmini Devi, M.E., Ph.D., Associate Professor, Computer Science and Engineering,** for her constant help and creative ideas over the periods of project work.

I am thankful to my **Faculty Members of Department of Computer Science and Engineering** for their continuous direction and guidance and timely support extended for me to carry out my project successfully.

Words are boundless to thank **Our Parents and Friends** for their constant encouragement to complete this project successfully.

DATE: 12.04.2023

[S.NAGA KANNIKA]

Programme: Master of Engineering- Computer Science and Engineering

Vision of the Department

To achieve education and research excellence in Computer Science and Engineering

Mission of the Department

M1: To excel in academic through effective teaching learning techniques

M2: To promote research in the area of computer science and engineering with the focus on innovation

M3: To transform students into technically competent professionals with societal and ethical responsibilities

Program Outcomes (POS)

PO1: An ability to independently carry out research / investigation and development work to solve practical problems.

PO2: An ability to write and present a substancial technical report/document.

PO3: Students should be able to demonstrate a degree of mastery over the area asper the specialization of the program .The mastery should be at a level higher than than the requirements in the appropriate bachelor program.

PO4: Ability to discriminate,evaluate,analyze and synthesize existing and new knowledge and integration of the same for enhancement of knowledge in Computer Science and Engineering.

PO5: Ability to think laterally and originally to identify,formulate and solve an engineering problem in Computer Science and Engineering and effectively utilize appropriate scientific and engineering techniques and methodologies in the problem solving process.

PO6: Ability to apply the tools from optimization, probability,statistics,simulation and engineering economic analysis,including fundamental application of the tools in IT industry invoving uncertainty and scarce or expensive resources.

Program Educational Objectives (PEOs)

PEO 1: To empower graduates to identify,create and solve computing problem by applying their knowledge of computing principles and mathematical theory to develop sustainable solutions to current and future computing problems.

PEO 2: To develop research attitude in graduates and to explore it for higher education Endeavors and constantly upgrade their skills with an attitude towards lifelong learning.

PEO 3: To facilitate graduates to acquire skills to communicate effectively with the society and contribute to the betterment of the society as a commited technical personnel.

ABSTRACT

Bitcoin uses a peer-to-peer technology to operate with no central authority or banks. Bitcoin is open-source; its design is public, nobody owns or controls Bitcoin and everyone can take part. Digital currency brings into use as open source software in pseudonymous creator Satoshi Nakamoto It is a crypto currency, so-called because it uses cryptography to control the creation and transfer of money. The goal of this work is to compare the accuracy of bitcoin price in USD prediction based on Long Short-term Memory (LSTM) network with self-attention. Real-time price data is collected by Pycurl from Bitfine. LSTM model is implemented by Keras and TensorFlow. The proposed model used in this work is mainly to present a classical comparison of time series forecasting, as expected, it could make efficient prediction limited in short-time interval, and the outcome depends on the time period. The LSTM could reach a better performance, with extra, indispensable time for model training, especially via CPU.

PROJECT MAPPED WITH PO AND PEO

ABSTRACT	PO's MAPPED	PEO's MAPPED
Bitcoin uses a peer-to-peer technology to operate with no central authority or banks. Bitcoin is open-source; its design is public, nobody owns or controls Bitcoin and everyone can take part. Digital currency brings into use as open source software in pseudonymous creator Satoshi Nakamoto It is a crypto currency, so-called because it uses cryptography to control the creation and transfer of money. The goal of this work is to compare the accuracy of bitcoin price in USD prediction based on Long Short-term Memory (LSTM) network with self-attention. Real-time price data is collected by Pycurl from Bitfine. LSTM model is implemented by Keras and TensorFlow. The proposed model used in this work is mainly to present a classical comparison of time series forecasting, as expected, it could make efficient prediction limited in short-time interval, and the outcome depends on the time period. The LSTM could reach a better performance, with extra, indispensable time for model training, especially via CPU.	PO1(L) PO2(M) PO3(H) PO4(M) PO5(M) PO6(M)	PEO1(M) PEO2(H) PEO3(M)

NOTE: L-LOW,M-MEDIUM,H-HIGH

12/9/23

SUPERVISOR

HEAD OF DEPARTMENT

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	i
	PROJECT MAPPED WITH PO AND PEO	ii
	TABLE OF CONTENTS	iii
	LIST OF TABLES	vi
	LIST OF FIGURES	vii
	LIST OF AND ABBREVIATIONS	ix
1	**INTRODUCTION**	1
	1.1 TRANSACTIONS	4
	1.2 OWNERSHIP	5
	1.3 MINING	6
	1.4 DECENTRALIZATION	7
	1.5 PRIVACY AND FUNGIBILITY	8
	1.6 HISTORY	8
	1.7 DATA MINING FOR BITCOIN PREDICTION	14
	1.8 DATA PREPROCESSING	14
	1.9 DATA MINING TECHNIQUES	15
	1.10 DATA MINING TASKS	16
	1.11 DATA MINING APPLICATIONS	18
	1.12 STATISTICAL AND DATAMININGTOOLS	18
2	**LITERATURE REVIEW**	21
3	**EXISTING SYSTEM**	30
	3.1 CLASSIFICATION AND PREDICTION ISSUES	30
	3.2 COMPARISON OF CLASSIFICATION AND PREDICTION METHODS	31
	3.3 TYPES OF ALGORITHMS IN DATA MINING	31
		32

CHAPTER NO	TITLE	PAGE NO
	3.3.2 MACHINE LEARNING-BASED APPROACH	32
	3.3.3 NEURAL NETWORK	32
	3.3.4 CLASSIFICATION ALGORITHMS IN DATA MINING	33
	3.3.5 ID3 ALGORITHM	33
	3.3.6 C4.5 ALGORITHM	34
	3.3.7 K NEAREST NEIGHBORS ALGORITHM	34
	3.3.8 NAÏVE BAYES ALGORITHM	35
	3.3.9 ANN ALGORITHM	36
	3.3.10 SUPPORT VECTOR MACHINES	37
4	**PROBLEM STATEMENT**	40
5	**PROPOSED SYSTEM**	41
	5.1 DATA SET COLLECTION	41
	5.2 PRE-PROCESSING	41
	5.3 TRAINING	41
	5.4 PREDICTION	41
	5.5 CONSTRUCTION OF ATTENTION-LSTM MODEL	42
	5.5.1 LSTM MODEL	42
	5.6 PROPOSED ALGORITHM 1	46
	5.7 EVALUATION	47
	5.7.1 CONFUSION MATRIX	47
	5.8 AREA UNDER CURVE	48

CHAPTER NO	TITLE	PAGE NO
6	**SOFTWARE SPECIFICATION**	50
	6.1 PYTHON	50
	6.1.1 WHY PYTHON?	50
	6.1.2 INSTALLATION	51
7	**SYSTEM IMPLEMENTATION**	52
	7.1 USE CASE	52
	7.2 SEQUENCE DIAGRAM	53
	7.3 ACTIVITY DIAGRAM	54
	7.1 USECASE	55
	7.5 CLASS DIAGRAM	56
8	**SIMULATION RESULT**	57
	8.1 DATASET AND PER PROCESSING	57
9	**CONCLUSION**	59
10	**FUTURE WORK**	60
	APPENDIX 1:SCREENSHOT	61
	APPENDIX 2:SOURCE CODE	66
	REFERENCES	70

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
2.1	Comparison Table for Literature Review	29

LIST OF FIGURES

FIGURE NO.	FIGURE NAME	PAGE NO.
1.1	Data structure of blocks in the ledger	2
1.2	Number of bitcoin transactions per month, semi logarithmic plot	2
1.3	Number of unspent transaction outputs	3
1.4	Bitcoin ownership	5
1.5	Bitcoin price	12
1.6	Data mining as a step in the process of knowledge discovery	14
1.7	Data mining tasks	16
1.8	Statistical and data mining tools	19
3.1	Date mining algorithms	31
5.1	Recurrent neural network (RNN)	42
5.2	The structure of LSTM cell	42
5.3	The process of adding an attention mechanism to the LSTM model	44
5.4	The internal structure of attention model	44
5.5	Attention-LSTM network architecture	46
5.6	Confusion Matrix	47
7.1	Use case	52
7.2	Sequence diagram	53
7.3	Activity diagram	54
7.4	Collaborative diagram	55
7.5	Class diagram	56
8.1	Data set description	57
8.2	High close plot	58
8.3	Predicted values	58
A.1	LSTM training	61
A.2	Confusion matrix	62
A.3	Precision score	62

FIGURE NO.	FIGURE NAME	PAGE NO.
A.4	Testing result	63
A.5	F-score analysis	64
A.6	Recall analysis	64
A.7	Precision analysis	65
A.8	Accuracy analysis	65

LISTS OF ABBREVIATIONS

SHA	-	Secure Hash Algorithm
POW	-	Proof Of Work
RPOW	-	Reusable Proof-Of-Work system
DB	-	Database
IRC		Internet Relay Chat
FINCEN	-	Financial Crimes Enforcement Network
MSB	-	Money Service Business
Seg Wit	-	Segregated Witness
CEC	-	Center For Cition Studies
KDD	-	Knowledge Discovery In Databases
SVM	-	Support Vector Machine
MAE	-	Mean Absolute Error
MSE	-	Mean Squared Error
RMSE	-	Roof Mean Squared Error
LSTM	-	Long-short Term Memory
ARIMAX	-	Auto Regressive Integrated Moving Average with exogenous input
BTC	-	Bitcoin
MBTC	-	Millibitcoin
RNN	-	Recurrent Neural Network
BNN	-	Bayesian Neural Network
MAPE	-	Mean Absolute Percentage
KNN	-	K Nearest Nearest Neighbour
ID3		Iterative Dichotomiser 3
ANN	-	Artificial Neural Networks
AUC	-	Area Under Curve
฿	-	Bitcoin
$	-	Dollar

AIRCELL BASED CONVOLUTIONAL NEURAL NETWORK TECHNIQUE FOR EGG QUALITY ANALYSIS

A PROJECT REPORT PHASE II

Submitted by

Mrs.G.RAMYA SHRI (927621MCS006)

in partial fulfillment for the award of the degree

of

MASTER OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

M.KUMARASAMY COLLEGE OF ENGINEERING

(An Autonomous Institution, Affiliated to Anna University, Chennai)

KARUR – 639 113

ANNA UNIVERSITY :: CHENNAI 600 025

APRIL 2023

M.KUMARASAMY COLLEGE OF ENGINEERING

(An Autonomous Institution, Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this project report "**AIRCELL BASED CONVOLUTIONAL NEURAL NETWORK TECHNIQUE FOR EGG QUALITY ANALYSIS**" is the bonafide work of "**G.RAMYASHRI (927621MCS006)**" who carried out the project work during the academic year 2022-2023 under my supervision.Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Signature

Dr. S.THILAGAMANI, M.E., Ph.D.,
SUPERVISOR,
Department of Computer Science and Engineering,
M.Kumarasamy College Of Engineering,
Thalavapalayam,Karur – 639 113.

Signature

Dr. M.MURUGASEN, M.E., Ph.D.,
HEAD OF THE DEPARTMENT,
Department of Computer Science and Engineering,
M.Kumarasamy College Of Engineering,
Thalavapalayam,Karur – 639 113.

Submitted for Project Work Phase II viva voce examination held on 12.04.2023

INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION

I affirm that the Project report titled " **AIRCELL BASED CONVOLUTIONAL NEURAL NETWORK TECHNIQUE FOR EGG QUALITY ANALYSIS** " being submitted in partial fulfillment for the award of Master of Engineering in Computer Science and Engineering, is the original work carried out by me. It has not formed the part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

G. Ramyashy

G.RAMYASHRI (927621MCS006)

I certify that the declaration made by the above candidate is true to the best of my knowledge.

S Thi 8/4/23

Name & Signature of the supervisor with date

Dr. S.THILAGAMANI, M.E.,Ph.d.
Professor & Head
Computer Science & Engineering
M. Kumarasamy College Of Engineering
Karur - 639 113

ACKNOWLEDGEMENT

Behind every achievement lies an unfathomable sea of gratitude to those who actuated it, without them it would have never came into existence, to them we lay the word of gratitude imprinted within us.

I would like to express my profuse gratitude to **Thiru.M.Kumarasamy, Chairman** and **Dr.K.Ramakrishnan, Secretary** of our college for providing extra ordinary infrastructure, which helped me in the completion of the project Phase II in time.

I wish to express my sincere thanks to our respected **Dr.B.S.Murugan, M.E., Ph.D., Principal,** for all the blessing and help provided during the period of project work.

I explore my gratitude to **Dr.M.Murugasen, M.E., Ph.D., Head of the Department, Computer Science and Engineering** for permitting me to undertake this project as part of the curriculum.

I am indebted to my project supervisor **Dr.S.Thilagamani ,M.E.,Ph.D., Professor and Dean, Computer Science and Engineering,** for her constant help and creative ideas over the period of project work .

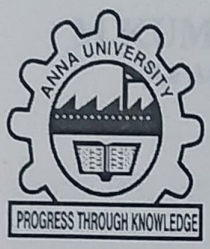
I am thankful to **my faculty members of CSE department** for their continuous direction and guidance and timely support extended for me to carry out my project successfully.

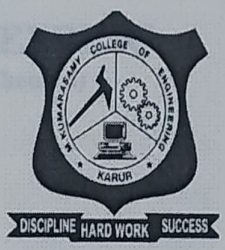
Words are boundless to thank **Our Parents and Friends** for their constant encouragement to complete this project successfully.

DATE: 8-4-2023

G. Ramyashy

[G.RAMYASHRI]





AN ANALYSIS OF DIFFERENT DATA SAMPLING METHOD FOR MEDICAL DATA USING MACHINE LEARNING

A PROJECT REPORT PHASE II

Submitted by

SHANMUGASUNDARAM P (927621MCS007)

in partial fulfillment for the award of the degree

of

MASTER OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

M.KUMARASAMY COLLEGE OF ENGINEERING
(An Autonomous Institution, Affiliated to Anna University, Chennai)
KARUR – 639 113

ANNA UNIVERSITY : CHENNAI 600 025

APRIL 2023

M.KUMARASAMY COLLEGE OF ENGINEERING

(An Autonomous Institution, Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this project report "**AN ANALYSIS OF DIFFERENT DATA SAMPLING METHOD FOR MEDICAL DATA USING MACHINE LEARNING**" is the bonafide work of "**SHANMUGASUNDARAM P (927621MCS007)**" who carried out the project work under my supervision. Certified further that to the best of knowledge the work reported herein does not form part of any other thesis or dissertion on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Signature
Mr.V. Mani, M.E .,
SUPERVISOR,
Department of Computer Science and Engineering,
M.Kumarasamy College Of Engineering,
Thalavapalayam,Karur – 639 113.

Signature
Dr. M.MURUGASEN, M.E., Ph.D.,
HEAD OF THE DEPARTMENT,
Department of Computer Science and Engineering,
M.Kumarasamy College Of Engineering,
Thalavapalayam,Karur – 639 113.

Submitted for Project Work Phase II viva voce examination held on 12/04/2023

INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION

I affirm that the Project report titled "**AN ANALYSIS OF DIFFERENT DATA SAMPLING METHOD FOR MEDICAL DATA USING MACHINE LEARNING**" being submitted in partial fulfillment for the award of **Master of Engineering in Computer Science and Engineering**, is the original work carried out by me. It has not formed the part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

P. Shanmugasundaram

SHANMUGASUNDARAM P (927621MCS007)

I certify that the declaration made by the above candidate is true to the best of my knowledge.

V. MANI & V.M 11/4/23

Name & Signature of the supervisor with date

ACKNOWLEDGEMENT

Behind every achievement lies an unfathomable sea of gratitude to those who actuated it, without them it would have never came into existence, to them we lay the word of gratitude imprinted within us.

I would like to express my profuse gratitude to **Thiru.M.Kumarasamy, Chairman** and **Dr.K.Ramakrishnan, Secretary** of our college for providing extra ordinary infrastructure, which helped me in the completion of the project Phase I in time.

It is a great privilege for us to express our gratitude to our esteemed **Principal Dr.B.S.Murugan , M.E., Ph.D.,** for providing us right ambiance for carrying out the project work.

I explore my gratitude to **Dr.M.MURUGESAN, M.E., Ph.D., Head, Department of Computer Science and Engineering** for his unwavering moral support throughout the evolution of the project.

I am indebted to my project supervisor **Mr.V.Mani, M.E., Associate Professor, Computer Science and Engineering,** for his constant help and creative ideas over the period of project work.

I am thankful to my faculty members of CSE department for their continuous direction and guidance and timely support extended for me to carry out my project successfully.

Words are boundless to thank **Our Parents and Friends** for their constant encouragement to complete this project successfully.

DATE: 12/04/2023

[SHANMUGASUNDARAM P]

M.KUMARASAMY COLLEGE OF ENGINEERING

Programme: Master of Engineering - Computer Science and Engineering

Vision of the Department

To achieve education and research excellence in Computer Science and Engineering

Mission of the Department

M1: To excel in academic through effective teaching learning techniques

M2: To promote research in the area of computer science and engineering with the focus on innovation

M3: To transform students into technically competent professionals with societal and ethical responsibilities

Program Outcomes (POS)

PO1: An ability to independently carry out research / investigation and development work to solve practical problems.

PO2: An ability to write and present a substancial technical report/document.

PO3: Students should be able to demonstrate a degree of mastery over the area asper the specialization of the program .The mastery should be at a level higher than than the requirements in the appropriate bachelor program.

PO4: Ability to discriminate,evaluate,analyze and synthesize existing and new knowledge and integration of the same for enhancement of knowledge in Computer Science and Engineering.

PO5: Ability to think laterally and originally to identify,formulate and solve an engineering problem in Computer Science and Engineering and effectively utilize appropriate scientific and engineering techniques and methodologies in the problem solving process.

PO6: Ability to apply the tools from optimization, probability,statistics,simulation and engineering economic analysis,including fundamental application of the tools in IT industry invoving uncertainty and scarce or expensive resources.

Program Educational Objectives (PEOs)

PEO 1: To empower graduates to identify,create and solve computing problem by applying their knowledge of computing principles and mathematical theory to develop sustainable solutions to current and future computing problems.

PEO 2: To develop research attitude in graduates and to explore it for higher education Endeavors and constantly upgrade their skills with an attitude towards lifelong learning.

PEO 3: To facilitate graduates to acquire skills to communicate effectively with the society and contribute to the betterment of the society as a commited technical personnel.

ABSTRACT

In today's era of internet, the amount of data generation is growing on increasing. With the development of artificial intelligence, big data classification technology provides the advantageous help for the medicine auxiliary diagnosis research. While due to the different conditions in the different sample collection, the medical big data is often imbalanced. Traditional classification algorithms usually assume that the number of samples in each class is similar and their misclassification cost during training is equal. However, the misclassification cost of patient samples is higher than that of healthy person samples. Therefore, how to increase the identification of patients without affecting the classification of healthy individuals is an urgent problem. The straightforward method to solve this problem is the resampling method by adding records to the minority class or deleting ones from the majority class. In this project, we have experimented with different SMOTE oversampling methods. The purpose of this work is to balance the imbalanced data using different sampling techniques. We use two medical dataset i.e PIMA INDIAN DIABETICS DATASET and CHRONIC KIDNEY DISEASE DATASET download from kaggle repository. So that in this project, split in to two phases. One is data sampling and other one is Prediction model. In this project we used different data sampling methods like SMOTE, K-means SMOTE, SMOTENC (SMOTE for Nominal and Continuous) and SMOTEN (SMOTE for Nominal). After getting modified data sampling dataset, to apply the different Machine learning algorithms i.e Decision tree, Random Forest, SVM and KNN to predict the prediction of Chronic Kidney Disease and diabetic disease in early stage. Based on Accuracy, Precision and Recall value from implemented tested machine learning model to find out the best Sampling as well as machine learning algorithms.

PROJECT MAPPED WITH PO AND PEO

ABSTRACT	PO's MAPPED	PEO's MAPPED
In today's era of internet, the amount of data generation is growing on increasing. With the development of artificial intelligence, big data classification technology provides the advantageous help for the medicine auxiliary diagnosis research. While due to the different conditions in the different sample collection, the medical big data is often imbalanced. Traditional classification algorithms usually assume that the number of samples in each class is similar and their misclassification cost during training is equal. However, the misclassification cost of patient samples is higher than that of healthy person samples. Therefore, how to increase the identification of patients without affecting the classification of healthy individuals is an urgent problem. The straightforward method to solve this problem is the resampling method by adding records to the minority class or deleting ones from the majority class. In this project, we have experimented with different SMOTE oversampling methods. The purpose of this work is to balance the imbalanced data using different sampling techniques. We use two medical dataset i.e PIMA INDIAN DIABETICS DATASET and CHRONIC KIDNEY DISEASE DATASET download from kaggle repository. So that in this project, split in to two phases. One is data sampling and other one is Prediction model. In this project we used different data sampling methods like SMOTE, K-means SMOTE, SMOTENC (SMOTE for Nominal and Continuous) and SMOTEN (SMOTE for Nominal). After getting modified data sampling dataset, to apply the different Machine learning algorithms i.e Decision tree, Random Forest, SVM and KNN to predict the prediction of Chronic Kidney Disease and diabetic disease in early stage. Based on Accuracy, Precision and Recall value from implemented tested machine learning model to find out the best Sampling as well as machine learning algorithms.	PO1(H) PO2(M) PO3(H) PO4(H) PO5(L) PO6(M)	PEO1(H) PEO2(M) PEO3(H)

NOTE: 1-LOW, 2-MEDIUM, 3-HIGH

V.m 11/4/23

SUPERVISOR

M. 7/11/23

HEAD OF THE DEPARTMENT

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	i
	PROJECT MAPPED WITH PO AND POE	ii
	TABLE OF CONTENTS	iii
	LIST OF TABLES	iv
	LIST OF FIGURES	v
	LIST OF ABBREVIATIONS	vi
1	INTRODUCTION	1
2	LITERATURE REVIEW	6
3	EXISTING SYSTEM	14
4	PROBLEM IDENTIFICATION	15
5	PROPOSED SYSTEM	17
6	SYSTEM REQUIREMENTS	18
7	METHEDOLOGY	19
8	IMPLEMENTATION	25
9	RESULT AND DISCUSSION	29
10	CONCLUSION AND FUTURE ENHANCEMENT	34
	APPENDIX 1 – SOURCE CODE	35
	APPENDIX 2 – SCREENSHOTS	50
	REFERENCES	55
	LIST OF PUBLICATION	59

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
2.1	COMPARISON TABLE FOR LITERATURE REVIEW	13
9.1	RESULT FOR ORIGINAL UNBALANCED DATASET	29
9.2	RESULT FOR SMOTE RESAMPLING DATA	29
9.3	RESULT FOR ADSAYN RESAMPLING DATA	30
9.4	RESULT FOR SMOTEOMEK RESAMPLING	30
9.5	RESULT FOR K-MEANS SMOTE RESAMPLING	30
9.6	DECISION TREE PERFORMANCE USING DIFFERENT DATA SAMPLING TECHNIQUES	31
9.7	KNN PERFORMANCE USING DIFFERENT DATA SAMPLING TECHNIQUES	31
9.8	SVM PERFORMANCE USING DIFFERENT DATA SAMPLING TECHNIQUES	32
9.9	RANDOM FOREST PERFORMANCE USING DIFFERENT DATA SAMPLING TECHNIQUES	33

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
1.1	DIFFERENT BETWEEN UNDERDAMPING AND OVERSAMPLING	2
1.2	CKD DISEASE DESCRIPTION	3
7.1	MACHINE LEARNING ALGORITHMS	21
8.1	DATA ARCHIECTURE DIAGRAM	25
8.2	WORKING PRINCIPAL OF SMOTE	26
8.3	WORKING PRINCIPAL OF ADAYEN	27
9.1	DECISION TREE PERFORMANCE GRAPH USING DIFFERENT DATA SAMPLING TECHNIQUES	31
9.2	KNN PERFORMANCE GRAPH USING DIFFERENT DATA SAMPLING TECHNIQUES	32
9.3	SVM DECISION TREE PERFORMANCE GRAPH USING DIFFERENT DATA SAMPLING TECHNIQUES	32
9.4	RANDOM FOREST PERFORMANCE GRAPH USING DIFFERENT DATA SAMPLING TECHNIQUES	33
A1	IMPORT LIBRARY AND DATASET	50
A2	CODE FOR GETTING CONFUSION MATRIX	50
A3	DECISION TREE IMPLEMENTATION	51
A4	KNN IMPLEMENTATION	51
A5	SVM IMPLEMENTATION	52
A6	RANDOM FOREST IMPLEMENTATION	52
A7	SMOTE IMPLEMENTATION	53
A8	ADASYN IMPLEMENTATION	53
A9	SMOTE + TOMEK IMPLEMENTATION	54
A10	K-MEANS SMOTE IMPLEMENTATION	54

LIST OF ABBREVIATIONS

ML	Machine Learning
CKD	Chronic Kidney Disease
SMOTE	Synthetic Minority Oversampling Technique
ADASYN	Adaptive Synthetic Sampling Approach
ACK	Acknowledgement
DT	Decision Tree
KNN	K- Nearest Neighbour
SVM	Support Vector Machine
RF	Random Forest
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
CSV	Comma Separated Value