

**A GBDT BASED SENTIMENT ANALYSIS AND
CLASSIFICATION OF TWITTER INFORMATION
USING OPTIMIZATION ALGORITHM**

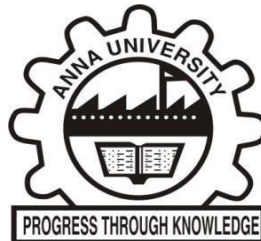
A THESIS

Submitted by

NEELAKANDAN S

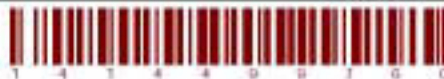
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY



**FACULTY OF INFORMATION AND COMMUNICATION
ENGINEERING
ANNA UNIVERSITY
CHENNAI 600 025**

MAY 2021



CERTIFICATE

This is to certify that all corrections and suggestions pointed out by the Indian /Foreign Examiner(s) are incorporated in the Thesis titled " A GBDT Based Sentiment Analysis and Classification of Twitter Information using Optimization Algorithm. " submitted by Mr. Neelakandan.S

Signature of the Supervisor

Dr.D Paulraj B.E., M.E., Ph.D.

Professor

Department of Computer Science and Engineering
R.M.D. Engineering College
Kavaraipett, Chennai - 601 206

Place :

Date :



CENTRE FOR RESEARCH

ANNA UNIVERSITY, CHENNAI-600 025



Proceedings of the Ph.D. Viva-Voce Examination of Mr.Neelakandan.S held at 11:00 AM on 07.05.2021 in Online Mode

The Ph.D. Viva-Voce Examination of Mr.Neelakandan.S (Reg. No. 1414499760) on his/her Ph.D. Thesis Entitled " A GBDT Based Sentiment Analysis and Classification of Twitter Information using Optimization Algorithm. " was conducted on **07.05.2021** at 11:00 AM in the Online Mode.

The following Members of the Oral Examination Board were present:

- | | |
|---|-----------------|
| 1. Dr. S Janakiraman, Assistant Professor, Department of Banking Technology School of Management, Pondicherry University, Kalapet Pondicherry - 605 014 | Indian Examiner |
| 2. Dr. B Ramadoss, Professor, Department of Computer Applications, National Institute of Technology, Tanjore Main Road NH 67 Near BHEL Thiruchirappalli - 620 015 | Subject Expert |
| 3. Dr. Paulraj.D, Professor, Department of Computer Science and Engineering, R.M.D. Engineering College, Chennai | Supervisor |


The research scholar, Mr. Neelakandan.S presented the salient features of his/her Ph.D. work. This was followed by questions from the board members. The questions raised by the Foreign and Indian Examiners were also put to the scholar. The scholar answered the questions to the full satisfaction of the board members.

The corrections suggested by the Indian/Foreign examiner have been carried out and incorporated in the Thesis before the Oral examination.

Based on the scholars research work, his/her presentation and also the clarifications and answers by the scholar to the questions, the board recommends that Mr.Neelakandan.S be awarded Ph.D. degree in the **Faculty of Information and Communication Engineering.**


Indian Examiner
[S. JANAKIRAMAN]


Subject Expert

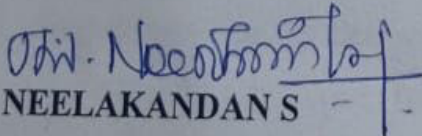

Supervisor
Dr.D.Paulraj B.E., M.E., Ph.D.
Professor
Department of Computer Science and Engineering
R.M.D. Engineering College
Kavaraipetti, Chennai - 601 206

**ANNA UNIVERSITY
CHENNAI 600 025**

BONAFIDE CERTIFICATE

The research work embodied in the present Thesis entitled
"A GBDT BASED SENTIMENT ANALYSIS AND CLASSIFICATION OF
TWITTER INFORMATION USING OPTIMIZATION ALGORITHM" has
been carried out in the Department of Computer Science and Engineering,
R.M.D. Engineering College, Kavaraipettai. The work reported herein is
original and 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 or to any other
scholar.

I understand the University's policy on plagiarism and declare that
the thesis and publications are my own work, except where specifically
acknowledged and has not been copied from other sources or been previously
submitted for award or assessment.


NEELAKANDAN S
RESEARCH SCHOLAR


Dr.D.PAULRAJ

SUPERVISOR

PROFESSOR

Department of Computer Science
and Engineering

R.M.D. Engineering College

Kavaraipettai - 601 206.

ACKNOWLEDGEMENT

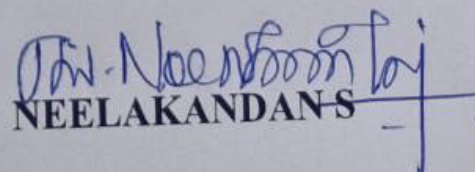
First, I would like to thank the almighty GOD for His blessings that have made everything possible by giving me the strength and courage to do this work.

I would like to express my sincere gratitude to my research supervisor **Dr.D.Paulraj**, Professor, Department of Computer Science and Engineering, R.M.D Engineering College, Kavaraipettai, Chennai, for his encouragement, motivation and invaluable guidance with my work during all stages, to bring this thesis into fulfilment.

I express my sincere thanks to my Doctoral Committee members, **Dr.R.Baskaran**, Associate Professor, Computer Science and Engineering, Anna University, CEG Campus, Chennai and **Dr.S.Godfrey Winstler**, Professor, Computer Science and Engineering, Saveetha Engineering College, Saveetha Nagar, Thandalam, Chennai for their constructive advice and valuable suggestions during various stages of this work.

I express my sincere thanks to the Head of the Department, R.M.D Engineering college, Kavaraipettai, Chennai, for all the support to research without any delay during my research period. I am grateful to **Dr.N.Marie Wilson**, Managing Director, Jeppiaar Institute of Technology, for his guidance and supporting me to carry out my research work. I extend my sincere thanks to my dear colleagues of Jeppiaar Institute of Technology for their support by providing me enough facilities to complete the research.

I wish to extend my special thanks to my family members, friends for their unstinted support and prayers. Finally, I thank and praise the lord almighty for providing me wisdom and strength to complete the research work successfully.


NEELAKANDANS

LIST OF TABLES

TABLE NO.	LIST OF TABLES	PAGE NO.
3.1	List of Sample Emoticons	83
3.2	Emoticon Features	91
3.3	Non-Emoticon Features	91
3.4	Precision for different techniques	94
4.1	Comparison of Existing techniques and the Proposed GBDT for 100 Numbers of Data	121
5.1	Details of parameters	130
5.2	Comparison of Precision Analysis - DLMNN	132
5.3	Comparison of Recall Analysis- DLMNN	134
5.4	Comparison of F-Measure Analysis - DLMNN	136
5.5	Comparison of Classification Accuracy - DLMNN	137
5.6	Comparison of Sentimental Score - DLMNN	139
5.7	Comparison of Time Complexity - DLMNN	140
5.8	Comparison of Precision Analysis -GBDT	142
5.9	Comparison of Recall Analysis -GBDT	144
5.10	Comparison of F-Measure -GBDT	145
5.11	Comparison of Classification Accuracy -GBDT	146
5.12	Comparison of Sentimental score -GBDT	148
5.13	Comparison of Time Complexity -GBDT	149

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
1.1	General Block Diagram of Sentiment Analysis Model	5
1.2	Block Diagram of Sentiment Analysis with Twitter Data	16
1.3	Process of Frequency Based Sentiment Mining	21
1.4	Process of Probability Based Sentiment Mining	22
1.5	Process of Relation Based Sentiment Mining	23
1.6	Process of Rule based sentiment mining	24
1.7	Process of pattern based sentiment mining	25
1.8	Architecture of NN Based Sentiment Analysis	27
3.1	Process of Tweet dataset and classification	73
3.2	System Architecture of Proposed DLMNN	80
3.3	Structure of DLMNN	85
3.4	Pseudo Code of PSO Algorithm	89
3.5	Comparative analysis of precision values	95
3.6	Recall for different techniques	96
3.7	F- score of various quantities of data in different methods	96
3.8	Time Complexity of different methods	97
3.9	Accuracy of the proposed and the existing systems	98
3.10	Comparative analysis of average sentiment score	99
4.1	The Proposed GBDT Framework	104
4.2	HDFS Map Reduce Framework	107

FIGURE NO.	TITLE	PAGE NO.
4.3	Pseudocode for I-EHO	114
4.4	Pseudocode for GBDT	120
4.5	Performance Analysis of Proposed GBDT in terms of Precision	122
4.6	Performance analysis of Proposed GBDT in terms of recall	123
4.7	Performance Analysis of proposed GBDT in terms of F-Score	124
4.8	Performance analysis of proposed GBDT in terms of Time Complexity.	125
4.9	Performance Analysis of proposed GBDT techniques in terms of Accuracy	126
4.10	Performance analysis of proposed GBDT in terms of Average Sentiment Score	127
5.1	Comparison of Precision Analysis - DLMNN	133
5.2	Comparison of Recall analysis- DLMNN	135
5.3	Comparison of F-Measure - DLMNN	137
5.4	Comparison of Classification Accuracy - DLMNN	138
5.5	Comparison of Sentimental Score - DLMNN	138
5.6	Comparison of Time Complexity - DLMNN	140
5.7	Performance Analysis of proposed GBDT and DLMNN in terms of Precision	142
5.8	Performance Analysis of proposed GBDT and DLMNN in terms of Recall	143
5.9	Performance Analysis of proposed GBDT and DLMNN in terms of F-Measure	145

FIGURE NO.	TITLE	PAGE NO.
5.10	Performance Analysis of proposed GBDT and DLMNN in terms of Accuracy	146
5.11	Sentimental Score Comparison of DLMNN & GBDT	147
5.12	Performance Analysis of proposed GBDT and DLMNN in terms of Time Complexity.	149
5.13	Performance Analysis of proposed GBDT and Existing models in terms of Precision, Recall, F-Score and Accuracy	150

LIST OF SYMBOLS AND ABBREVIATION

API	-	Application Programming Interface
ANN	-	Artificial Neural Network
BVA	-	Block Based Visual Analysis
CFS	-	Candidate Feature Selection
CCA	-	Canonical Correlation Analysis
CNN	-	Convolutional Neural Network
CISER	-	Credibility, Interest, And Sentiment Enhanced Recommendation
CRM	-	Customer Relationship Management
DT	-	Decision Tree
DL	-	Deep Learning
DLMNN	-	Deep Learning Modified Neural Network
DSA	-	Dual Sentiment Analysis
DBN	-	Dynamic Bayesian Network
EHO	-	Elephant Herd Optimization
ET	-	Extend-Type
FFNN	-	Feed Forward Neural Network
GFS	-	Google File System
GBDT	-	Gradient Boost Decision Tree
GRN	-	Graph Regression Networks
HDFS	-	Hadoop Distributed File System
HLDA	-	Hierarchical Latent Dirichlet Allocation
I-EHO	-	Improved Elephant Herd Optimization
IDF	-	Inverse Document Frequency values
KFCV	-	K – Fold Cross Validation
KLP	-	Knowledge Learning Process
LDA	-	Linear Discriminant Analysis

LSTMs	-	Long Term Short-Term Memory
ML	-	Machine Learning
NLP	-	Natural Language Processing
NN	-	Neural Network
POS	-	Part Of Speech
PSO	-	Particle Swarm Optimization
PSO	-	Particle Swarm Optimization
PCA	-	Principal Component Analysis
RNN	-	Retinal Neural Networks
SA	-	Sentiment Analysis
SC	-	Sentiment Classification
SDAFI	-	Social Data Analysis And Feedback Information
SDA	-	Spatial Data Analysis
SSTL	-	Stanford Sense Tree Library
SML	-	Supervised Machine Learning
SVM	-	Support Vector Machine
TF	-	Term Frequency
TF-IDF	-	Term Frequency-Inverse Document Frequency
WOA	-	Whale Optimization Algorithm