

FAKE DATA ANALYSIS AND DETECTION USING ENSEMBLED HYBRID ALGORITHM

A PROJECT REPORT

Submitted by

PALAGATI BHANU PRAKASH REDDY (0015113135)

MANDI PAVAN KUMAR REDDY (0015113134)

GANJIKUNTA MANASWINI REDDY (0015113144)

Under the guidance of

Dr. K. M. Mehata, Senior Professor, CSE Department

*in partial fulfillment for the award of the degree
of*

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING



HINDUSTAN
INSTITUTE OF TECHNOLOGY & SCIENCE
(DEEMED TO BE UNIVERSITY)

**HINDUSTAN INSTITUTE OF TECHNOLOGY AND
SCIENCE, CHENNAI - 603 103**

APRIL 2019



HINDUSTAN
INSTITUTE OF TECHNOLOGY & SCIENCE
(DEEMED TO BE UNIVERSITY)

BONAFIDE CERTIFICATE

Certified that this project report **“FAKE DATA ANALYSIS AND DETECTION USING ENSEMBLED HYBRID ALGORITHM”** is the bonafide work of **“PALAGATI BHANU PRAKASH REDDY (0015113135) MANDI PAVAN KUMAR REDDY (0015113134) GANJIKUNTA MANASWINI REDDY (0015113144)”** who carried out the project work under my supervision during the academic year **2018-2019**.

SIGNATURE

SIGNATURE

HEAD OF THE DEPARTMENT

SUPERVISOR

Dr. Rajeswari Mukesh

Dr. K. M. Mehata

Professor

Senior Professor

Department of Computer Science
and Engineering

Department of Computer Science
and Engineering

Hindustan Institute of
Technology and Science

Hindustan Institute of
Technology and Science

INTERNAL EXAMINER

EXTERNAL EXAMINER

Name: _____

Name: _____

Designation: _____

Designation: _____

Institution Name: _____

Project Viva - voce conducted on _____

ACKNOWLEDGEMENT

It's our extreme pleasure to thank our Chancellor **Dr. (Mrs.) Elizabeth Verghese** and Vice Chancellor **Dr. K. P. Issac**, Hindustan Institute of Technology and Science for providing conducive environment which helped us to pursue our project in a diligent and an efficient manner.

We wish to express our sincere gratitude to Dean (Academics) **Dr. N. Vasudevan**, Hindustan Institute of Technology and Science for his valuable directions, suggestions and support.

We are thankful to **Dr. M. Rajeswari Mukesh**, Head of the Department, Electronics and Communication Engineering for having evinced keen interest in our project and for his continued support.

We are indebted to our project guide **Dr. K. M. Mehata**, Senior Professor, Electronics and Communication Engineering for his/her valuable guidance and technical support in the accomplishment of our project.

We also thank our Project Coordinators **Mr. S. K. Shankar**, Associate Professor, Computer Science and Engineering for their support and coordination throughout the accomplishment of the project.

Our sincere thanks to all the teaching and non-teaching staff and family members who have been constantly supporting us throughout the accomplishment of this project.

BHANU PRAKASH REDDY. P

PAVAN KUMAR REDDY. M

MANASWINI REDDY. G

DEDICATION

This project “**FAKE DATA ANALYSIS AND DETECTION USING ENSEMBLED HYBRID ALGORITHM**”, dedicated to our beloved parents and friends. This project is also dedicated to the staffs of Department of Computer Science and Engineering, Hindustan Institute of Technology and Science.

ABSTRACT

Fake data detection is the most important problem to be addressed in the recent years, there is lot of research going on in this field. Because of its serious impacts on the readers. Researchers, government and private agencies working together to solve the issue. This project represents a hybrid approach for fake data detection using the multinomial voting algorithm. This algorithm was tested with multiple fake news dataset which resulted in an accuracy score of 94 percent which is a benchmark in the field of machine learning where the other algorithms are at a range of 82 to 88 percent. The list of algorithms that have been used here is as follows Naïve Bayes, Random Forest, Decision Tree, Support Vector Machine, K Nearest Neighbours. All these algorithms use training data as the bag of words model which was created using Count Vectorizer. Experimental data has collected from the Kaggle data world. Python is used as a language to verify and validate the results. Tableau is used as a visualization tool. Implementation is carried out using default algorithm values. A fake news detection website is created to validate and visualize the real time use cases of the algorithm. This will create a change if it is used properly.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ACKNOWLEDGEMENT	iii
	ABSTRACT	v
	LIST OF FIGURES	x
1.	INTRODUCTION	1
	1.1 Introduction	2
	1.2 About	3
	1.3 Summary	3
2.	LITERATURE REVIEW	4
	2.1 Literature Review	5
	2.2 Summary	6
3.	PROJECT OVERVIEW	7
	3.1 Project Overview	8
	3.2 Hardware Requirements	8
	3.3 Software Requirements	8
	3.4 Block Diagram Description	9
	3.4.1 Website	9
	3.4.2 Login	10
	3.4.3 Testing	10
	3.4.4 Algorithm	10
	3.4.5 Prediction	11
	3.4.6 Output	11
	3.4.7 Database	11
	3.4.8 History	12

	3.4.9 User Access	12
	3.4.10 Admin Access	12
4.	ALGORITHM ANALYSIS	13
	4.1 Introduction	14
	4.2 Fake Data Properties	14
	4.3 Training Testing Data	14
	4.4 Analyzer Definitions	15
	4.5 Naïve Bayes Algorithm	16
	4.5.1 Confusion Matrix Visualization	17
	4.6 SVM for Fake Data Analysis	19
	4.6.1 Confusion Matrix Visualization	20
	4.7 Random Forest for Fake Data Analysis	21
	4.7.1 Confusion Matrix Visualization	22
	4.8 KNN for Fake Data Analysis	23
	4.8.1 Confusion Matrix Visualization	24
	4.9 Decision Tree for Fake Data Analysis	25
	4.9.1 Confusion Matrix Visualization	27
	4.10 Implementation of the Ensembled Algorithm	28
	4.10.1 Algorithm Explanation	30
	4.10.2 Confusion Matrix Visualization	33
	4.10.3 Performance Analysis of Proposed Algorithm	34
	4.11 Summary	37
5.	SYSTEM DESIGN	38
	5.1 Introduction	39

5.2 Objectives of thee Design	39
5.3 Factors considered in the design	39
5.4 Output Design	40
5.5 Input Design	40
5.6 Home Page	41
5.7 Register	41
5.8 Login Page	42
5.9 About Us	43
5.10 Search Page	44
5.11 Result Page	44
5.12 History	45
5.13 Edit Profile	46
5.14 Change Password	47
5.15 Admin View	47
5.16 Logout	48
5.17 System Architecture	49
5.18 Database Connectivity	50
5.19 Summary	51
6. TESTING	52
6.1 Introduction	53
6.2 Testing Methods	54
6.2.1 Performance Testing	54
6.2.2 Black Box Testing	54
6.2.3 Unit Testing	54
6.2.4 Selenium Testing	54

	6.2.5 Python Testing	55
	6.3 Summary	55
7.	CONCLUSION AND FUTURE WORK	57
	7.1 Conclusion	58
	7.2 Future Work	59
8.	REFERENCES	60
9	SAMPLE CODE	64

LIST OF FIGURES

FIG NO.	NAME OF THE FIGURE	PAGE NO.
3.4	Block diagram of Fake Data Analysis website	9
3.4.7	Database architecture diagram of the project	11
4.5.1.1	Pie chart representation of Naïve Bayes confusion matrix	18
4.5.1.2	Naïve Bayes analysis line chart	18
4.6.1.1	Pie chart representation of SVM confusion matrix	20
4.6.1.2	SVM analysis line chart	21
4.7.1.1	Pie chart representation of Random Forest confusion matrix	22
4.7.1.2	Random Forest analysis line chart	23
4.8.1.1	Pie chart representation of KNN confusion matrix	24
4.8.1.2	KNN analysis line chart	25
4.9.1.1	Pie chart representation of KNN confusion matrix	27
4.9.1.2	Decision Tree analysis line chart	28
4.10	Hybrid algorithm architecture diagram	30
4.10.1.1	Pie chart representation of Hybrid Ensembled algorithm confusion matrix	33
4.10.1.2	Hybrid Ensembled algorithm analysis line chart	34
4.10.3.1	Overall Comparisons	36
4.10.3.2	Comparison of Accuracy's	37
5.6	Homepage	41
5.7	Register	42
5.8	Login Page	42
5.9	About Us	43
5.10	Search Page	44

5.11	Result Page	45
5.12	History	46
5.13	Edit Profile	46
5.14	Change Password	47
5.15	Admin View	48
5.16	Logout	48
5.17	System Architecture	49
5.18	Database Connectivity	50
6.2.4	Selenium test result for the web app software	55