PROJECT REPORT

On

SPAM DETECTION USING BIG DATA & MACHINE LEARNING

Submitted to Rajasthan Technical University
in partial fulfillment of the requirement for the award of the degree of

B.TECH.

in

COMPUTER ENGINEERING

Submitted By
Aayushi Bhatt (PIET15CE003)
Vibhor Bhimsariya (PIET15CE118)
Unnati Singhal (PIET15CE115)

Under the Guidance of
Mr. Deepak Moud

at

POORNIMA INSTITUTE OF ENGINEERING & TECHNOLOGY,
JAIPUR
RAJASTHAN TECHNICAL UNIVERSITY, KOTA
APRIL, 2019

CERTIFICATE

This is to be certified that the project entitled "**SPAM DETECTION USING BIG DATA AND MACHINE LEARNING** " has been submitted for the Bachelor of Computer Science and Engineering, Poornima Institute Of Engineering & Technology, Jaipur during the academic year 2018-2019 is a Bonafede piece of project work carried out by " **Aayushi Bhatt, Vibhor Bhimsariya & Unnati Singhal** " towards the partial fulfillment for the award of the Degree (B.Tech.) under the guidance of "**Mr. Deepak Moud**" and supervision and no part of thereof has been submitted by them for any degree or diploma.

Project Guide	Project Coordinator	Mr. Deepak Moud
Mr. Deepak Moud	Prof. (Dr.) Praveen Gupta	(H.O.D., CSE)
(Assistant Professor)	(Professor)	

CANDIDATE'S DECLARATION

We, **Aayushi Bhatt (PIET15CE003), Vibhor Bhimsariya (PIET15CE118) & Unnati Singhal(PIET15CE115)** B. Tech (Semester- VIII) of "**Poornima Institute Of Engineering & Technology, Jaipur",** hereby declare that the Project Report entitled **"SPAM DETECTION USING BIG DATA AND MACHINE LEARNING"** is an original work and data provided in the study is authentic to the best of our knowledge. This report has not been submitted to any other Institute for the award of any other degree.

AAYUSHI BHATT	**VIBHOR BHIMSARIYA**	**UNNATI SINGHAL**
PIET15CE003	**PIET15CE118**	**PIET15CE115**

Place: **Jaipur**

Date: **23-10-2018**

ACKNOWLEDGEMENT

It is our pleasure to be indebted to various people, who directly or indirectly contributed in the development of this work and who influenced our thinking, behavior and acts during the course of study.

We express our sincere gratitude to ***Dr. O. P. Sharma,*** Director, PIET for providing us an opportunity to undergo this Major Project as the part of the curriculum.

We are thankful to ***Mr. Deepak Moud, HOD, CS*** for his support, cooperation, and motivation provided to us during the training for constant inspiration, presence and blessings.

We are thankful to ***Mr. Puneet Mathur*** for his support, cooperation, and motivation provided to us during the training for constant inspiration, presence and blessings.

We also extend our sincere appreciation to ***Prof. (Dr.) Praveen Gupta*** who provided his valuable suggestions and precious time in accomplishing our Project report.

Lastly, we would like to thank the almighty and our parents for their moral support and friends with whom we shared our day-to-day experience and received lots of suggestions that improved our quality of work.

AAYUSHI BHATT	**VIBHOR BHIMSARIYA**	**UNNATI SINGHAL**
PIET15CE003	**PIET15CE118**	**PIET15CE115**

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

E-Mail is one of the well-known communication services in which a message is send electronically. The maximization of the digital use by different organizations has prompted the expanded utilization of Email Services. This ascent pulled in assailants, which have brought about E-Mail Spam problem. Spam messages include advertisements, free services, promotions, awards, etc. People are using the ubiquity of mobile phone devices is expanding day by day as they give a vast variety of services by reducing the cost of services and maximization of Digital Service. E-Mail is one of the broadly utilized communication service. In any case, this has prompted an expansion in E-Mail attacks like E-Mail Spam.

In this problem, preliminary results are mentioned or explained herein based on publicly available datasets. This problem is further expanded using multiple background datasets.

Spam filtering is the process of detecting the unwanted or unsolicited email or text from getting into the user's inbox. Spam filtering applications work on text filters. Text filters work by using algorithms to detect which words and phrases are most often used in the spam emails. we will build two spam classifications engine one by using logistic regression and the other by Naive Bayes. Finally, we will check the accuracy of these engines by using Machine Learning and Big Data

KEYWORDS: Hadoop Distributed File System, MapReduce, Spark, Naive Bayes, Logistic Regression