TABLE OF CONTENTS

Title	Page No.
ACKNOWLEDGEMENT	vi
ABSTRACT	vii
TABLE OF CONTENTS	viii-ix
CHAPTER 1 INTRODUCTION	1
1.1 Data Growth	2-3
1.2 Text Classification	3
1.3 Problem Statement	3
1.4 Motivation	4
1.5 Contribution	4
CHAPTER 2 LITERATURE SURVEY	5-6
CHAPTER 3 METHODOLOGY	7-10
3.1 Dataset	11
3.2 Preprocessing	11-12
3.3 Feature Engineering	13
3.3.1 TF-IDF	13-14
3.3.1.1 TF-IDF (n-grams)	14
3.3.2 CountVectorizer	14-15
3.3.2.1 CountVectorizer (n-gram)	15
3.4 Doc2Vec	15
3.5 Labelling	15
3.6 Another Approach	16
3.7 Algorithm	16
3.8 Classification	17

3.8.1 Classifiers 17-19
3.9 Evaluation Matric 19-20

CHAPTER 4 TOOLS 21
4.1 Python Programming Language 21
4.1.1 Python Over Other Programming Language 21
4.1.2 Python In Machine Learning 21
4.1.3 Python In Natural Language Processing 22
4.2 Libraries 22

CHAPTER 5 RESULT AND DISCUSSION
5.1 Result 24-65
5.2 With Clustering (ongoing development) 65-66
5.3 Discussion 66-69

CHAPTER 6 CONCLUSION 70
6.1 Future Scope 70-71

References 72-73