

TABLE OF CONTENT

PREFACE	i
TABLE OF CONTENT	ii
TABLE OF FIGURES	v
LIST OF TABLES	vii
ABSTRACT.....	viii
INTISARI.....	ix
CHAPTER I INTRODUCTION	1
1.1 Research background	1
1.2 Research Problem.....	2
1.3 Research Objectives	2
1.4 Research Scope.....	2
1.5 Research Benefits	3
CHAPTER II LITERATURE REVIEW	4
CHAPTER III BASIC THEORY.....	22
3.1 Weak Supervision.....	22
3.1.1 <i>Rubrix</i> framework	23
3.1.2 Uploading dataset into <i>Rubrix</i> framework	23
3.1.3 Adding rulesets in <i>Rubrix</i> framework	24
3.1.4 Weak supervision performance metrics	25
3.1.5 Fitting label model with Snorkel.....	26
3.2 Pre-processing	27
3.3 LSTM	28
3.4 Word Embedding	29
3.5 SVM	29

3.6	Performance metrics (for Classification).....	30
3.6.1	Confusion matrix.....	30
3.6.2	Accuracy	30
3.6.3	Precision.....	31
3.6.4	Recall.....	31
3.6.5	F1-Score	31
CHAPTER IV RESEARCH METHODOLOGY		32
4.1	Research description.....	32
4.2	Tools and materials	35
4.2.1	Tools.....	35
4.2.2	Material	36
4.3	Data collection and Dataset Description	36
4.4	Dataset preparation.....	38
4.5	Ruleset writing using <i>Rubrix</i> and label modeling	40
4.6	Classifier training/learning	42
4.7	Performance comparison / evaluation	43
4.8	Data Mixing.....	43
CHAPTER V IMPLEMENTATION.....		44
5.1	Dataset Collection	44
5.1.1	Dataset Scrapping for Weak Supervision	44
5.1.2	Dataset Preparation for Supervised Learning	46
5.2	Weak Labeling Dataset for Weak Supervision	46
5.2.1	Importing Dataset to <i>Rubrix</i>	46
5.2.2	Query-based Label Modeling.....	48
5.2.3	Evaluate label model	51