

TABLE OF CONTENTS

APPROVAL PAGE	I
TABLE OF CONTENTS	II
LIST OF TABLES	IV
LIST OF FIGURES	V
ABSTRACT	VI
CHAPTER I. INTRODUCTION	1
1.1 Research background	1
1.2 Research problem	3
1.3 Research scope	4
1.4 Research objective	4
1.5 Research advantages	4
CHAPTER II. LITERATURE REVIEW	6
2.1 Overview of Graph Machine Learning	6
2.2 Conventional vs Graph Machine Learning	7
2.3 Overview of Graph Convolutional Networks	8
2.4 Evaluation of Graph Convolutional Networks	10
2.5 Conclusion and Needs for Additional Research	12
CHAPTER III. THEORETICAL BASIS	13
3.1 Graph Theory	18
3.2 Machine Learning	20
3.3 Graph Machine Learning	21
3.3.1 Graph Representation Learning	22
3.3.2 Graph Neural Network	24
3.3.2.1 Graph Convolutional Networks	24
3.4 PinSAGE	25
3.5 Performance evaluation	31
3.5.1 Confusion matrix	31
3.5.2 Accuracy	32
3.5.3 Recall	32
3.5.4 Precision	33
3.5.5 F-beta score	33

CHAPTER IV. RESEARCH METHODOLOGY	35
4.1 Research description	35
4.2 Tools	35
4.3 Dataset	36
4.4 Experiment workflow	38
4.5 Data collection	39
4.6 Data preparation	39
4.7 Data preprocessing	40
4.8 Exploratory data analysis	42
4.9 Feature engineering and selection	42
4.10 Data conversion	45
4.11 Representation learning	47
4.12 Model training	47
4.13 Hyperparameter tuning	49
4.14 Model evaluation	50
CHAPTER V. IMPLEMENTATION	51
5.1 Data collection implementation	51
5.2 Data preparation implementation	51
5.3 Data preprocessing implementation	53
5.3.1 Dropping redundant columns	57
5.3.2 Handle column values	59
5.3.3 Preprocess continuous column values	60
5.4 Exploratory data analysis implementation	63
5.5 Feature engineering and selection implementation	66
5.5.1 Feature engineering	66
5.5.2 Encode categorical features	67
5.5.3 Selecting important features	72
5.6 Data conversion implementation	73
5.6.1 Node and edge feature imputation	74
5.6.2. Finalizing and saving the graph	75
5.6.3 Final dataset statistics	76
5.7 Model setup	77



5.7.1 PinSAGE model class definition	77
5.7.2 PinSAGE model training function definition	78
5.7.3 Hyperparameter tuning setup	79
5.8 Model training	80
5.9 Model testing	81
5.10 Data evaluation	81
CHAPTER VI. RESULT AND DISCUSSION	85
6.1 Result	85
6.1.1 PinSAGE results	86
6.1.2 BGNN results	88
6.1.3 RGCN results	90
6.2 Discussion	91
CHAPTER VII. CONCLUSION AND FUTURE WORKS	93
7.1 Conclusion	93
7.2 Future Works	93
REFERENCES	95