1 Experiment 1: Before Perturbation (Experiment on Feature Data without Removal Edges)

DESCRIPTION: Experiment I is conducted on statistical tests between original data and normalized feature data.

1. Subsection 1.1 explains statistical tests for classification based the original feature compare to normalized feature. The original feature consists of Netpro2VecMetgraphs, Netpro2VecNDD, Netpro2VecTM1, Netpro2VecTM2, Netpro2VecNDD+TM1, Netpro2VecNDD+TM1+TM2, Graph2Vec, GL2Vec, FeatherGraph, and SF as our graph embeddings schemes. Meanwhile, normalized feature data refers to all original features data that have been normalized using the min-max method, ensuring that feature values fall within the interval [0,1].
2. In this experiment, we implement both groups of graph embeddings in classification models and evaluate their performance through statistical tests, including accuracy, precision, recall, and F1-score from all models-based graph embeddings.
3. Following Table 1a until Table 1d ilustrate the test results.
4. Section 1.2 presents Figure 1a until Figure 1d as visualization of the statistical test results

RESUME: The comparison resume is based on the percentage difference of the original features compare to the normalized features. Rely on the accuracy, precision, recall, and F1-score, we describe the best and the lowest percentage difference as follow:

1. The best percentage difference of accuracies are in GCNs, DT, GNB, and RF with percentage difference of original compare to their normalized feature, that is about 0.000%. Meanwhile the lowest is in non linear SVM (SVMnl), that is about 0.051%.
2. The best percentage difference of precision is in DT, that is about 0.000%, meanwhile the lowest is in SVMnl, that is about 0.047%.
3. The best recall are in GCNs and DT, that is about 0.000%, meanwhile the lowest is in SVMnl, that is about 0.051%.
4. The best percentage difference of F1-score are in GCNs and DT, that is about 0.000%, meanwhile the lowest score is in SVMnl, that is about 0.069%.

CONCLUSION: Based on the statistical tests,analyzing the effect of normalized features compared to the original features, we conclude that the best-performing model is DT, achieving 0.000% accuracy, precision, recall, and F1-score. This is followed by GCNs, which also maintain 0.000% accuracy, recall, and F1-score, except for precision, which shows a slight variation of 0.018%.

1.1 Classification based Original Vs Normalized Features

Table 1: **Comparison between Original and Normalized Features**

(a) **Accuracy**

Accuracy

Rank	Models:	Original	Normalized	Percentage diff.
1	GCNs	1	1	0,000
2	SVML	0.97	0.98	0,010
3	KNN	0.965	0.968	0,003
4	DT	0.948	0.948	0,000
5	GNB	0.935	0.935	0,000
6	RF	0.93	0.93	0,000
7	SVMnl	0.883	0.928	0,051

(b) **Precision**

Precision

Rank	Models:	Original	Normalized	Percentage diff.
1	GCNs	0.982	1	0,018
2	SVML	0.972	0,982	0,010
3	KNN	0.952	0.972	0,021
4	GNB/SVMnl	0.951	0.951	0,000
5	RF/GNB	0.945	0.948	0,003
6	RF	0.933	0.945	0,013
7	SVMnl/RF	0.891	0.933	0,047

(c) **Recall**

Recall

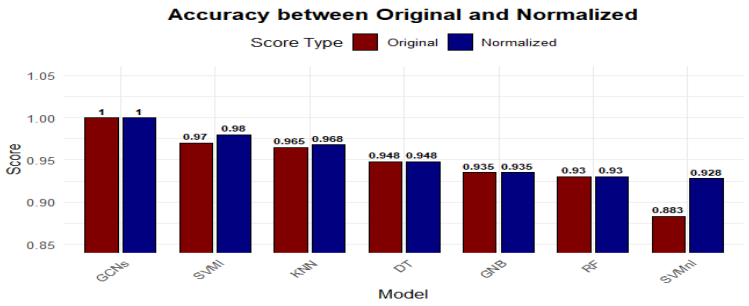
Rank	Models:	Original	Normalized	Percentage diff.
1	GCNs	1	1	0,000
2	KNN	0.965	0.979	0,015
3	SVML	0.949	0.967	0,019
4	DT	0.947	0.947	0.000
5	GNB	0.933	0.945	0,013
6	RF	0.93	0.933	0,003
7	SVMnl	0.855	0.93	0,051

(d) **F1-Score**

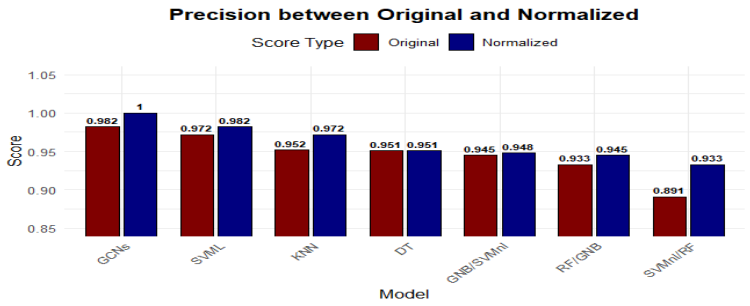
F1-Score

Rank	Models:	Original	Normalized	Percentage diff.
1	GCNs	1	1	0.000
2	SVML	0,969	0,98	0,011
3	KNN	0,965	0,967	0,002
4	DT	0,947	0,947	0,000
5	GNB	0,932	0,945	0,014
6	RF	0,929	0,932	0,003
7	SVMNL	0,869	0,929	0,069

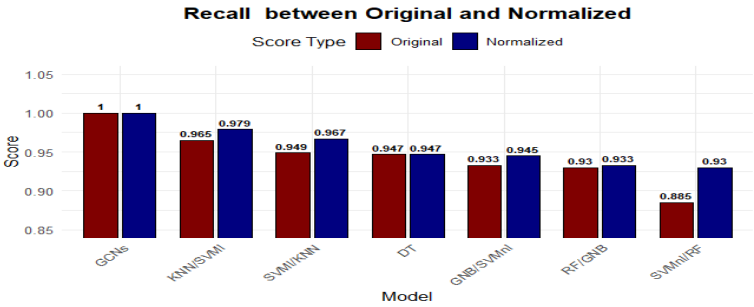
1.2 Figure of Statistical Tests Result



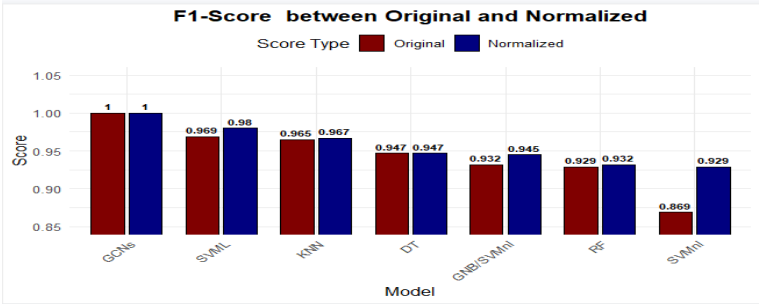
(a) Accuracy



(b) Precision



(c) Recall



(d) F1-Score

Figure 1: Plot between Original and Normalized Feature