# Introduction

# Related work

## Causality introduction

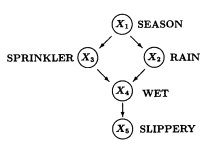
1. DAG

Figure1

Form1:

Form2:

1. SEM…

Form3:

Form4:

Form5：

Form6:

## Student performance prediction

1. Feature selection-relevance

[ref2]

1. Causality based approaches

# Students' performance prediction based on causal inference and multi-head self-attention mechanism

## Factor-association discovery using causal analysis

1. Notears algorithm to reveal linear SEM\

Form7:

Form8:

Form9:

Form10:

Form11:

Form12:

Form13:

(a) if and only if is acyclic (i.e. );  
(b) The values of quantify the "DAG-ness" of the graph;  
(c) is smooth;  
(d) and its derivatives are easy to compute.

Form14:

,.

Form15:

for all and all

Form16:

form17:

Form18:

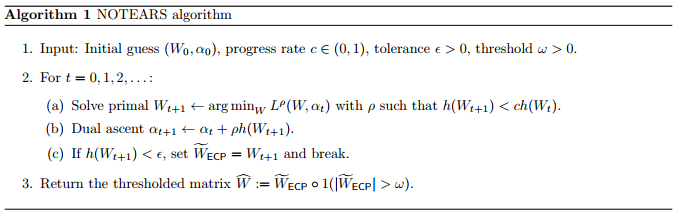
Form19:

Form20:

Form21:

Form22:

Form23:



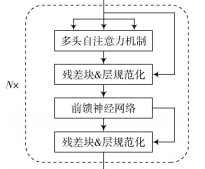
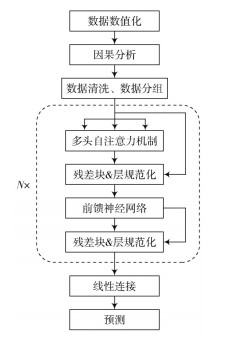
Form24:

**Here,**

**Form25:**

Form26:

## Transformer



# Result

## Dataset

## Assessment metrics

Form27:

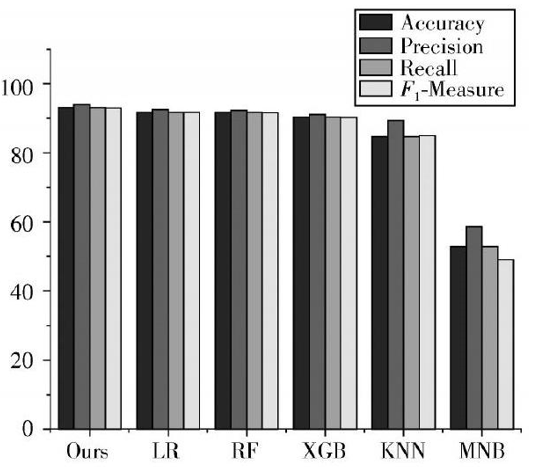
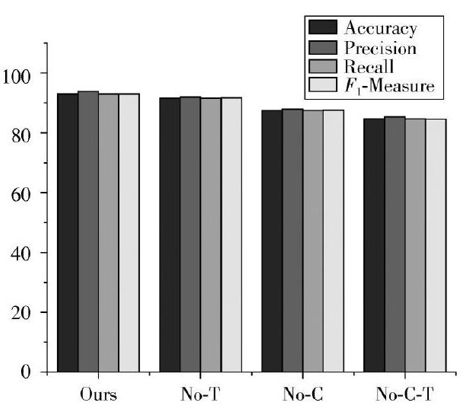
## Parameters

## result

## prediction performance

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 方法 | Accuracy | Precision | Recall | -Measure |
| Ours | 93.06 | 93.90 | 93.06 | 93.01 |
| LR | 91.67 | 92.38 | 91.67 | 91.65 |
| RF | 91.67 | 92.17 | 91.67 | 91.63 |
| XGB | 90.28 | 91.10 | 90.28 | 90.17 |
| KNN | 84.72 | 89.34 | 84.72 | 85.03 |
| MNB | 52.78 | 58.60 | 52.78 | 49.11 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 方法 | Accuracy | Precision | Recall | -Measure |
| Ours | 93.06 | 93.90 | 93.06 | 93.01 |
| No-T | 91.67 | 91.92 | 91.67 | 91.71 |
| No-C | 87.50 | 87.87 | 87.50 | 87.57 |
| No-C-T | 84.72 | 85.38 | 84.72 | 84.57 |



# Ref

[1]Judea Pearl, causality-models, reasoning and inference, 2009, Cambridge

[2]Kimcholmyong

[3] POLYZOU A, KARYPIS G. Feature extraction for next-term prediction of poor student performance [J]. IEEE transactions on learning technologies, 2019, 12(2): 237-248.  
[4] HUANG S, FANG N. Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models [J]. Computers & education, 2013, 61: 133-145.

[5] PANDEY M, SHARMA V K. A decision tree algorithm pertaining to the student performance analysis and prediction [J]. International journal of computer applications, 2013, 61(13): 1-5.

[6] 谢娟英, 张宜, 陈恩红. 学生成绩关键因素挖掘与成绩预测[J].南京信息工程大学学报(自然科学版), 2022,11(3):316-325.

[7] MA Y, CUI C, NIE X, et al. Pre-course student performance prediction with multi-instance multi-label learning [J]. Science China information sciences, 2019, 62(2): 1-3.

[8] 李梦莹, 王晓东,院书岚, 等.基于双路注意力机制的学生成绩预测模型[J].计算机研究与发展, 2022, 57(8):1729-1740.

[9]Rijisong

[10] A. Nemirovski. Optimization II: Standard Numerical Methods for Nonlinear Continuous Optimization. 1999

[11] K. Zhong, I. E.-H. Yen, I. S. Dhillon, and P. K. Ravikumar. Proximal quasi-newton for computationally

intensive l1-regularized m-estimators. In *Advances in Neural Information Processing Systems*, pages

2375{2383,2014

# Summary