# Project Overview

Highlight the contribution of this work. Including: existing related studies, the improvement/the contribution of this project.

Briefly explain the results of project in the current stage.

# Project System and Model Design

## Project Architecture Design Overview

1. System and Mode Architecture Figure (please show how to combine clustering and recommendation model together)
2. Description of this design. What’s the innovation of this design.

## Data Overview [data for both clustering and recommendation model]

1. Data schema, data source,

|  |  |  |  |
| --- | --- | --- | --- |
|  | Name | Description | Data Amount |
| VM Demographics information | VM-subnet &sg mapping Index | The log file from … that records the info about …. | 25K… |
|  | …. |  |  |
|  | …. |  |  |
|  |  |  |  |
| VM Communication Behavior | VM communication ip pairs |  |  |
|  | …. |  |  |

1. Data preprocessing processes/tools/methods
   1. Log scale --- details
   2. Pca/…. – details
   3. ….
   4. Ip matrix ---- represent the dynamic VM-VM interaction activities

show the schema of ip matrix

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Src 1 | … | Src 2 | … | Src 3 |  |
| Vm1 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

1. Dataset overview (distribution of data, features… )
   1. Label 1: --- describe the process/rules for labeling , what does it mean in the project working scenario, how many labels under this rule
   2. Label 2:---
   3. Label 3:---

….

## System and Model Design

1. Clustering
   1. Algorithms --- details
      1. Algorithm 1
      2. Algorithm 2
   2. Parameter setting -- details
   3. Merge --- details
      1. Algorithm
      2. different threshold settings
2. Recommendation model
   1. Key components—details with mathematic equations
      1. Feature extraction
         1. Process
         2. Reference
         3. Data input/output in this stage
      2. feature embedding,
      3. attention layers,
      4. convolutional layers
      5. softmax layer

## Result

1. Experiment data structure
2. x\_i <f\_i, k\_i, g\_i…>
3. f\_i: explain what does this item for? In our project, why do we consider this one a one of the key features of input for this model.
4. K\_i:….
5. ….
6. y\_i <…..>
   * 1. label1: …how and why to label in this way? What does it represent for in this case.
     2. Label 2: …
     3. ….
7. Experiment processes
   1. Clustering parameter setting/ merge threshold… etc.
   2. Recommendation model training processes, parameter settings…
      1. Training
      2. Validation
      3. Testing
8. Result analysis

Clustering output figures

Thresholds

Figures

Clustering result performance table

Model final output:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Threshold | Label1 | Precision | Recall | F1 | Total recommended ips | Total paired ips in ground truth |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Threshold | Label2 | Precision | Recall | F1 | Total recommended ips | Total paired ips in ground truth |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |