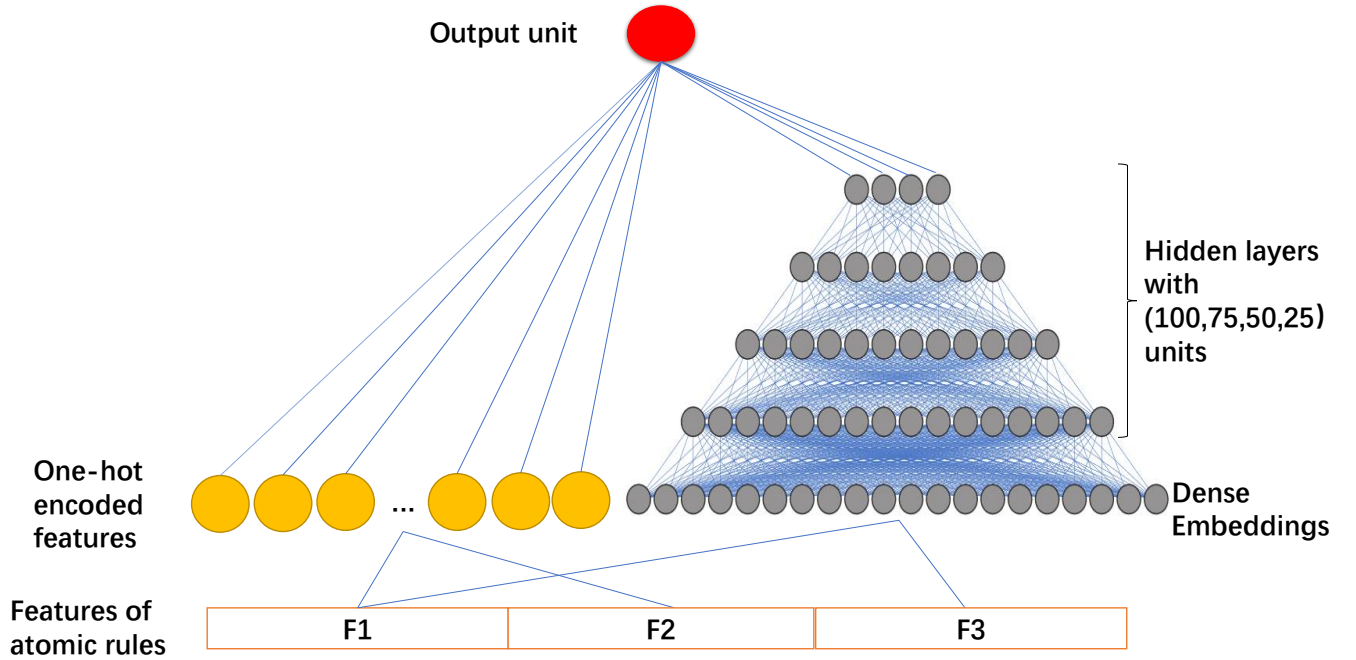# 1 OVERVIEW OF THE MACHINE LEARNING MODEL

The model we use consists of a logistic regression and a DNN, and its structure is shown in Figure 1. In particular, there are four fully connected layers in the DNN, of which the size is 100, 75, 50, 25, respectively. Additionally, Feature F1 (the basic semantics in atomic rules) and F2 (the user ID of runing processes) are fed into the logistic regression model, F1 and F3 (the rule comments) are fed into the DNN. Table 1 list the input layers for our model.

Table 1: Features of the Model

Feature Types	Feature Name	Description
Base Columns	subject	the subject field in a rule
	object	the object field in a rule
	class	the class field in a rule
	permission	the permission field in a rule
	domain	whether the subject is a process or not
	mlstrusted	whether the subject is a mlstrusted process or not (allows to read up and write down)
	core	whether the subject is a vendor-specified process or not
	app	whether the subject is an app process or not
	net	whether the subject is an app process with network process
	user	the range of user id
Crossed Columns	(object,class)	class is used for polishing the type of object
	(class,permission)	the permission is confined by class
	(object, class, permission)	The tuple indicates the behavior of a subject
	(subject, domain, mlstrusted, core, app, net)	This tuple constitutes the attribute vector of a subject
Deep Columns	64 dimensional embedding of subject	the embedding vector representing subject
	64 dimensional embedding of object	the embedding vector representing object
	8 dimensional embedding of class	the embedding vector representing class
	8 dimensional embedding of permission	the embedding vector representing permission
	300 dimensional embedding of allow comments	the embedding vector representing allow comments of the subject
	300 dimensional embedding of neverallow comments	the embedding vector representing neverallow comments of the subject

Figure 1: Overview of the Model