



**K. N. Toosi University of Technology**

**Faculty of Physics  
Educational Group of  
Atomic-Molecular and Astronomy**

# **Machine Learning Projects (Project 1)**

by  
**Ali Bagheri**

**Teacher**  
**Dr. Mohammad Hossein Zhoolideh**

**Academic Year 1401-1402  
(First Semester)**

## Diagnosis of Diabetes

In this project, we ask you to predict whether a patient has diabetes or not based on the measurements in a dataset. You can download the dataset needed to answer this question from this [link](#).<sup>1</sup>

There are several variables in this dataset: a number of independent variables along with a target dependent variable. Each of these variables is as follows:

Columns	Description
Pregnancies	number of pregnancies
Glucose	glucose level in blood
Blood Pressure	blood pressure measurement
Skin Thickness	thickness of the skin
Insulin	insulin level in blood
BMI	Body Mass Index
Diabetes Pedigree Function	diabetes percentage
Age	age of the patient
Outcome	final result 1 is Yes and 0 is No

**Note:** The given data is raw. To answer this question, you must first preprocess the data using the Pandas package.

## Important Points

Be sure to

- Leave appropriate comments for different parts of your code.
- Completely explain about the algorithm(s) you use to answer this question.
- Measure your model performance using model evaluation metrics and interpret the obtained result(s).

**A part of your score will be allocated to these items.**

\* You should write all the steps of your project in the **Jupyter notebook** and upload it as a file with the **.ipynb** extension on the vc site.

---

<sup>1</sup>To save the dataset, you need to press Ctrl+S on the opened page and save the .csv file