

**Ahsanullah University of Science and Technology**

*Department of Computer Science & Engineering*

|  |  |
| --- | --- |
| Course No. | CSE 4108 |
| Course Name | Artificial Intelligence Lab |
| Project | E-commerce customer's purchase prediction using their profile data |

**Submitted To:**

Md. Siam Ansary Tonmoy Hossain

Department of CSE, AUST Department of CSE, AUST

**Submitted By:**

|  |  |
| --- | --- |
| Name | Easinur Rashid Prantick |
| ID No. | 17.02.04.030 |
| Session | Fall - 2020 |
| Section | A (A2) |

**Documentation of the Dataset**

1. **Size of the Dataset:**
   1. Number of columns: 5.

- User\_ID

- Age

- Gender

- P\_Code

- C\_Order

* 1. Number of Rows: 118.

1. **Description of the features with their units:**

E-commerce marketing is getting intelligent day to day using their user’s data. In this data set there are 5 Colum named User\_ID, Age, Gender, P\_Code, C\_Order. Here P\_code is Product ID and C\_Order is conformation of an order.

* User\_ID (independent variable,X1): [data type-integer]
  + User ID number
* Age (independent variable,X2): [data type-integer]
  + People of various age limit
* Gender (independent variable,X3): [data type-integer]
  + Male (1)
  + Female (0)
* P\_Code (independent variable,X4): [data type-integer]
  + Product ID number
* C\_Order (dependent variable,Y): [data type-integer]
  + Confirmed an order (1)
  + Cancel an order (0)

In any site, a gender of user order a product. But the product depends on the taste and the age of a user, gender too. We stored which gender ordered which product depending on their age and gender. The C\_0rder is 1 when the user confirmed any order. Depending on the accuracy of the dataset every user will see related products depending on their interest individually.

1. **Citations of the websites:** N/A (self-made)
2. **Distribution of the labels:**

In dependent variable column(Y), There are 71 “0” labelled data which means in the Dependent column 71 user didn’t confirm the order. And rest of 47 “1” labelled data which means in the Dependent column 47 user confirm the order.

So here, 60% of the user cancel a confirmed order and 40% of the user finally confirm the purchase.