**Assignment – 3**

***Social Data Mining Techniques***

***Topic: -***

* **Technology**

**Dataset:**

Source:

<https://www.kaggle.com/datasets/jmmvutu/dating-app-lovoo-user-profiles>

**PART A: Draft a Decision Tree Analysis**

**Steps and Working:** Decision Tree is a learning technique used for both classification and regression models. It breaks down the structure in the form of nodes and leaves. Root is the first node and leaves is terminal node.

**Read CSV**: It will the load and read the dataset which is CSV format. Here I have loaded the dataset contains user profile information of users of the website *Lovoo*.

**Store:** This will store the data in repository. This will help next operator to fetch from it. Here I have saved it in the name data in assignment 3 repository.

Retrieve: Retrieve is used to fetch the data stored in store. This will retrieve the data for further procedures of decision tree.

**Set Role:** It is used to set the attribute to target variable. Here gender looking is the targeted variable and assign target role as label. Then decision tree is categorised by gender looking.

**Filter Examples:** This will filter out all the examples in the dataset. Here for both gender looking and isonline category, I have given the filter as not missing. This will display the data with no missing value.

Decision Tree: This operator build tree like structure from the input given. That is, it will display the decision tree according to gender looking with maximum depth of 10.

***Process: -***

A screenshot of a computer

Description automatically generated

***Output: -***

1. Filtered Data

A screenshot of a computer

Description automatically generated

b) **Decision Tree:** with nodes and leaves as gender looking. As we can see most females are looking male as their matches. But in some cases females are looking for both genders and female partners.

A screenshot of a computer

Description automatically generated

c) **Description**

Graphical user interface, text, application

Description automatically generated

**Part B: Logistic regression Analysis technique**

**Read EXCEL**: It will the load and read the dataset which is Excel format. Here I have loaded the data of dating app user’s profile.

**Store:** This will store the data in repository. This will help next operator to fetch from it

Retrieve: Retrieve is used to fetch the data stored in store. This will retrieve the dating app data for further procedures of decision tree.

**Select Attribute:** what this operator does is it selects a particular attribute in which users is interested in and add filters to those attributes only like comparing. My filter attribute type is subset and attributes are age, verified and counts pictures.

**Numerical to Binominal: -** This operator changes the type of the selected numeric attributes to a binominal type. It also maps all values of these attributes to corresponding binominal values.

Set Role: It is used to set the attribute to target variable. Here, my attributes filter type is single and attributes verified.

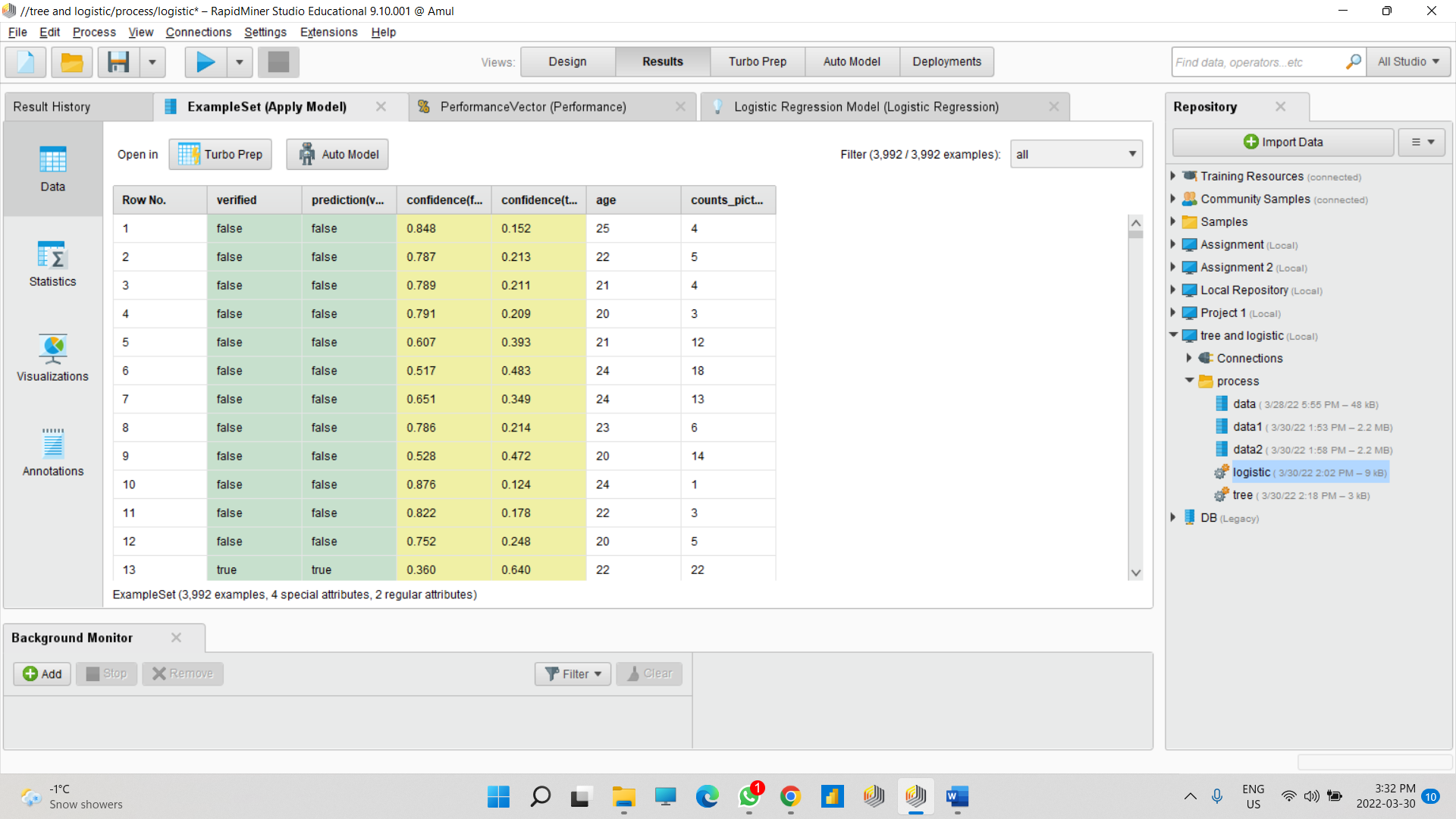
**Logistic Regression: -** This operator is a simplified version of the Generalized Linear Model operator. To perform Logistic Regression, the Family parameter is set automatically to binomial, and the link parameter to logit. Here, my solver is set at AUTO.

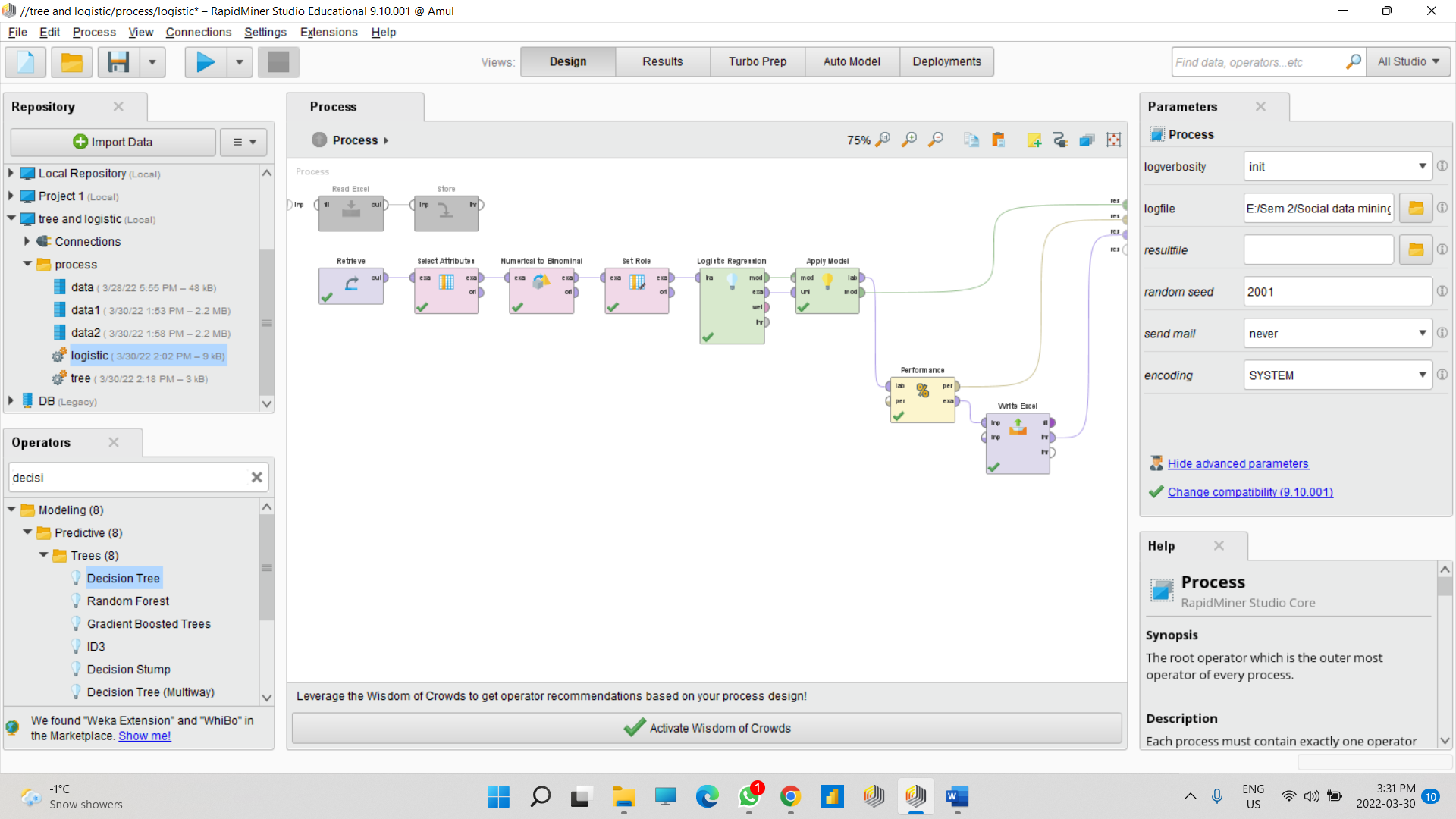
**Apply model: -** This operator applies a model on an example set.

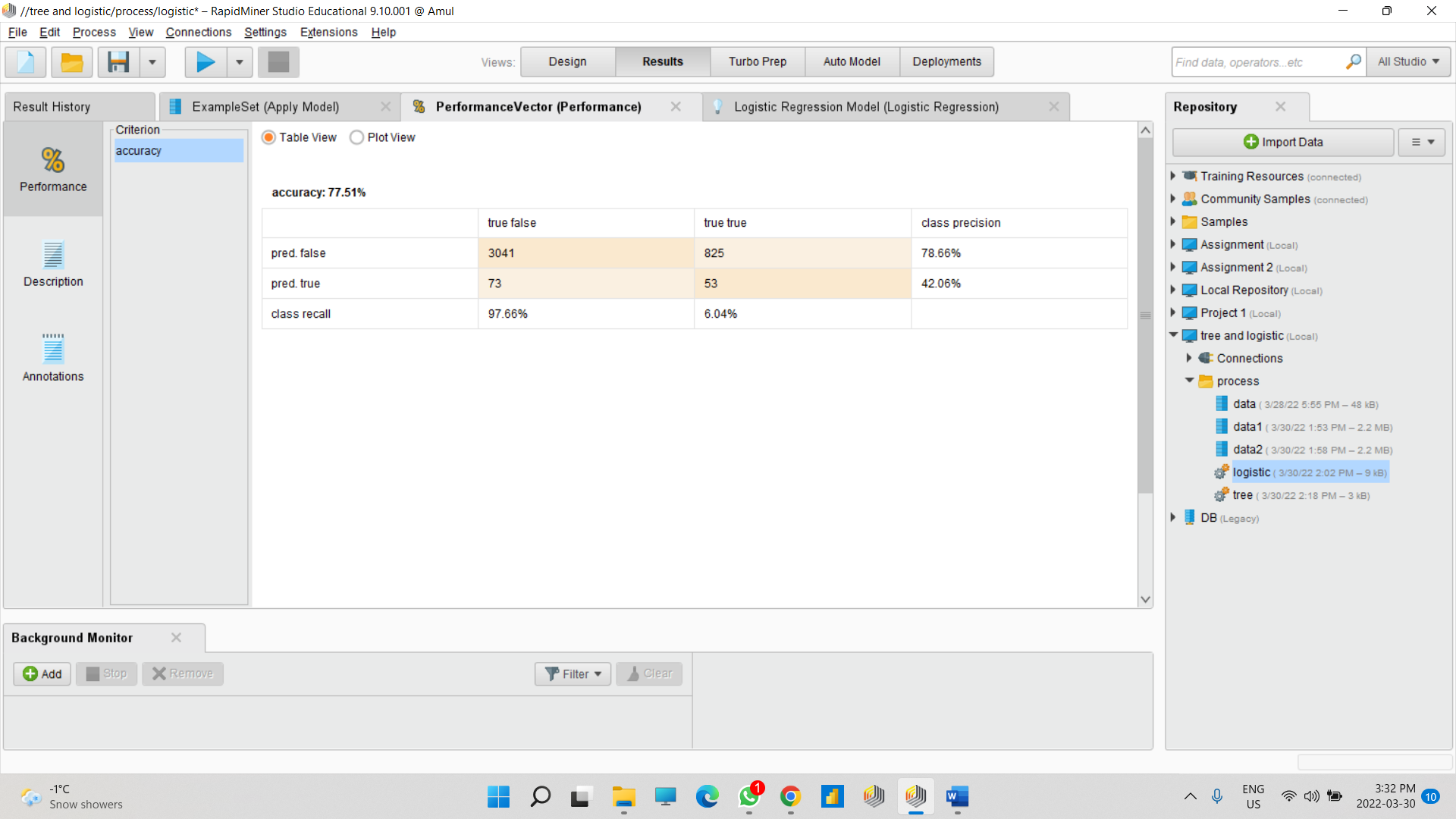
**Performance: -** This operator is used for statistical performance evaluation of classification tasks. This operator delivers a list of performance criteria values of the classification task.

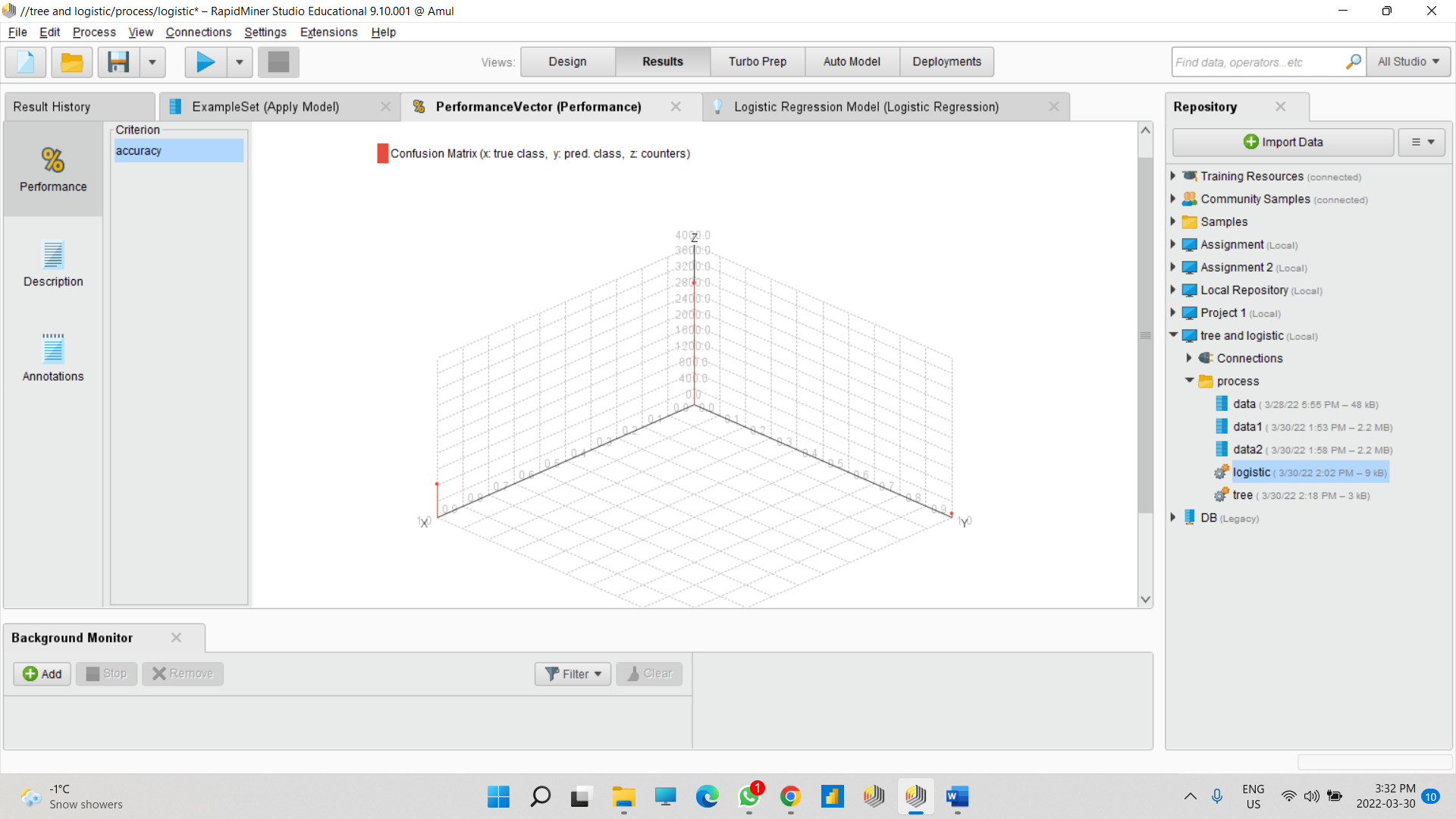
**Write excel: -** This operator writes Example Sets to an Excel spreadsheet file.

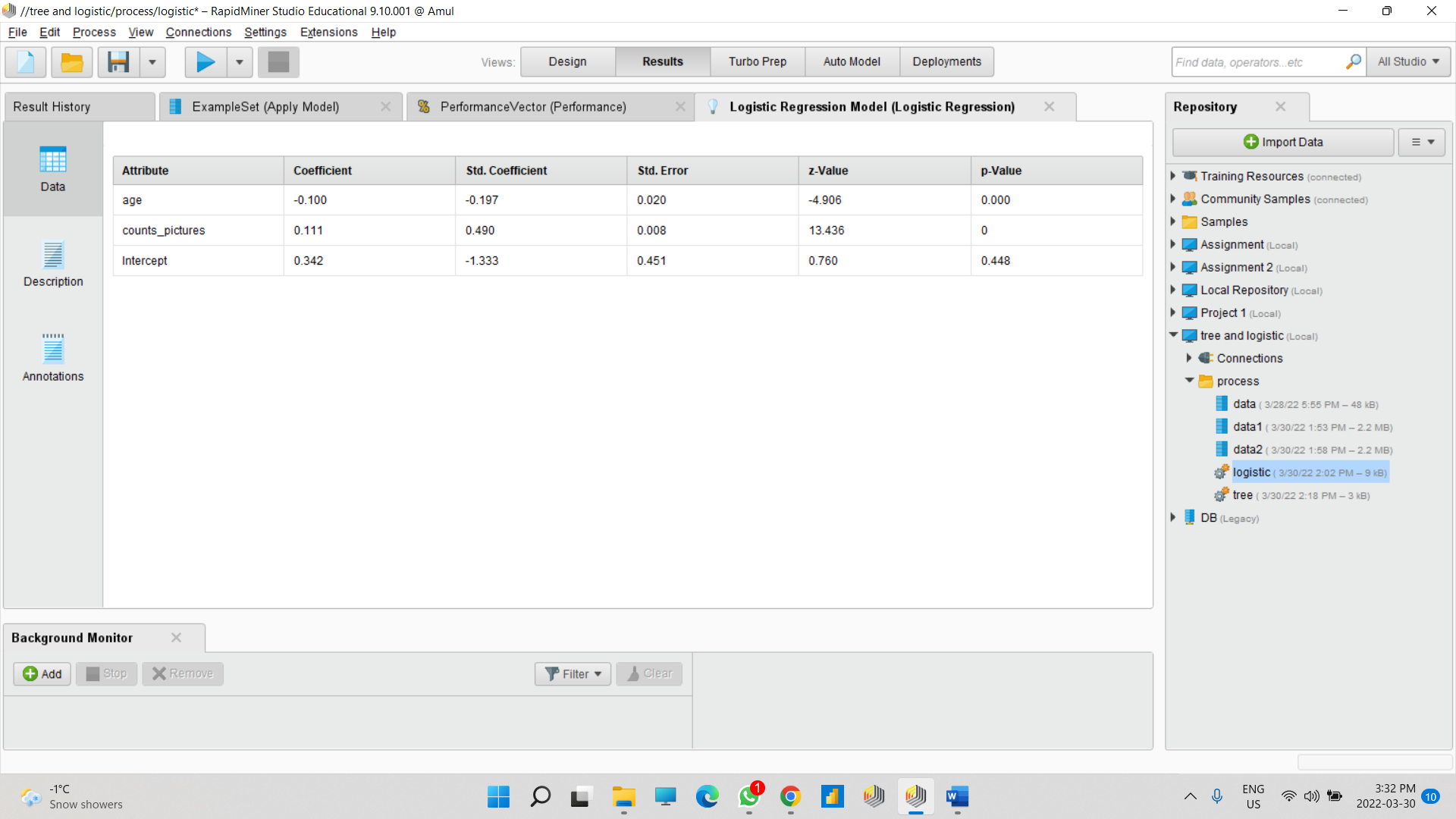
**Outputs:**











**Part D : Explain the following: Under what circumstances would you adopt Logistic regression Analysis Technique and what the visualization chart /diagram demonstrates.**

**About the dataset:**

This dataset contains user profile information of females from social network called Lovoo. The data from this dating app helps to gather better user profile. The dataset was collected during spring 2015 (April, may).

**Reason for choosing Dataset:**

We had to perform particular task like the Decision Tress making and for that we needing attributes like gender as male and female to categorize the data. And to perform logistic regression we needed one dependent binary variable and one or more nominal independent variable and this was the perfect dataset for our task.

**Explanation:**

Logistic Regression Technique is a type of classification technique used in Machine Learning. It is a statistical method used to predict outcomes of dependent variable based on previous observations. It is a kind of analysis which is commonly used algorithm for solving binary classifications problems. It is applied to predict the categorical dependent variable. It is used when the prediction is categorical, or example, Yes or No, True or False, 0 or 1. In this predicted output regression can either be one of them, no middle grounds.

Visualization depicts the relationship between the dependent variable and one or more independent variables. The logistic regression plot shows that the probability of success drops as the temperature rises. The slope of the line is not very steep when the temperatures in the data are near 50 indicating that the likelihood falls slowly as the temperature rises.